(9)

AN INTERACTIVE NEURAL NETWORK SYSTEM FOR ACOUSTIC SIGNAL CLASSIFICATION

CONTRACT NO. N00014-89-C-0237

FINAL REPORT

SUBMITTED TO:

Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217-5000

SUBMITTED BY:

Advanced Resource Development Corporation
9151 Rumsey Road
Columbia, Maryland 21045

February 28, 1990

AN INTERACTIVE NEURAL NETWORK SYSTEM FOR ACOUSTIC SIGNAL CLASSIFICATION

CONTRACT NO. N00014-89-C-0237

FINAL REPORT

SUBMITTED TO:

Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217-5000

SUBMITTED BY:

Advanced Resource Development Corporation
9151 Rumsey Road
Columbia, Maryland 21045

February 28, 1990

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704 0188

1a REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS N. A.
2a SECURITY CLASSIFICATION AUTHORITY N. A.	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b DECLASSIFICATION/DOWNGRADING SCHEDULE N. A.	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report ONR-89-1	5 MONITORING ORGANIZATION REPORT NUMBER(S) Same
6a NAME OF PERFORMING ORGANIZATION Advanced Resource Development Corporation 6b OFFICE SYMBOL (If applicable)	7a NAME OF MONITORING ORGANIZATION Office of Naval Research
6c ADDRESS (City, State, and ZIP Code) 9195 Rumsey Road Columbia, MD 21045	7b ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, VA 22217-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research 8b OFFICE SYMBOL (If applicable) Code 1142PS	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-89-C-0237
8c ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, VA 22217-5000	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
61153N 42	RR040209	RR04020901	400o041sbi0

11 TITLE (Include Security Classification)
(U) An Interactive Neural Network System For Acoustic Signal Classification

12 PERSONAL AUTHOR(S)
Nelson Steel, Michael Philips, Georgianna Meagher

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final	FROM 89/09/01 TO 90/02/28	90/02/28	123

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	An Interactive Neural Network System for Acoustic Signal Classification

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The objective of this project was to develop an understanding of the effect of neural networks, implemented in interactive systems, on sonar operators and other naval personnel. Specifically, the project called for the development of a prototype system, employing neural networks to test the effect of interactive (man-in-the-loop) operations. ARD developed such a system which is used to classify time domain signals generated from the insonification of an underwater mine-like target. The system converts the time domain signals to frequency domain and frequency over time (spectrograms) and displays the signals at the users' request in all three formats. A time windowing function is also provided to allow the user to closely inspect specific portions of the time domain signal. In addition, a neural network system classifies the signal according to three parameters: shell thickness, interior content and angle of

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL John J. O'Hare	22b TELEPHONE (Include Area Code) (202) 696-4502 22c OFFICE SYMBOL Code 1142PS

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

insonification. The operator has the option to use any or all of the available tools to classify the signal. Specific performance was measured over a series of experiments to determine the type of interactions, tool usage, dependence on the neural networks and other parameters. Results have shown that most users exhibit a large bias towards the use of the neural network analysis because of their highly accurate classification. Those users who independently do well in classifying signals tend not to depend on the aid of the networks when working with clean signals. All users have a bias towards using the networks when the signals are corrupted with noise. Future work will concentrate on the integration of neural network tools into existing systems in real-world situations. A better understanding of the human-network interactions will be gained when the ability of the networks to classify real world signals is decreased due to the complex geometries of actual mines and environmental effects on the sonar returns (thermoclines, shallow water, surface returns).

TABLE OF CONTENTS

ARD

ARD

LIST OF FIGURES

1.0 OVERVIEW

1.1 Introduction

The Office of Naval Research is interested in the interaction of human operators with neural networks or connectionist-based systems when trying to determine the source of an acoustic signal. ARD was awarded a Phase I SBIR contract in September 1989 to develop an acoustic classification system employing an array of traditional and network-based tools to be used for an analysis of the type mentioned above. This report will describe the foundation on which the classification system was developed, the system itself, the experiments conducted and the results. It should be noted that ARD has completed all the work outlined in the Phase I proposal and is prepared to use the results discussed in this report to move into a Phase II effort. The findings to date show a definite bias towards the use of "perfect classifiers" in the type of experimentation conducted in this project. This will need to be further analyzed in a more realistic environment using real world signals and existing classification systems to better determine the true effect of integrating neural network classification systems into the decision-making process. An approach to how this can be tested is described in ARD's Phase II proposal.

1.2 Rationale and Approach to the Phase I Effort

The use of acoustic sensors for the automatic detection and classification of underwater objects such as mines is of considerable importance. Ideally, backscattered active sonar returns from remote objects would be processed automatically to determine their composition, orientation, contents, and other characteristics. Although it is well known that the sonar return contains a great deal of information about the physical properties of the insonified object (Hickling, 1962; Morse, 1983), it can be extremely difficult to exploit this information for practical use. Analytic solutions have been derived to calculate the pressure field for idealized spherical or cylindrical objects, and numerical methods can extend these solutions to more complex geometries (Stanton, 1989). Nevertheless, most real world objects are too complex to permit a detailed theoretical or numerical analysis of their sound scattering properties.

Despite this, everyday experience suggests that acoustic classification of this sort is possible. For example, human listeners distinguish the sounds of "wooden" from "metallic" or "solid" from "hollow" objects with surprising ease. Perceptual psychologists have investigated the ability of human listeners to identify the source events for a wide range of environmental sounds. These include machinery noise (Talamo, 1982), the sounds of metallic (Howard, 1983) and non-metallic impacts (Warren & Verbrugge, 1984), classroom sounds (Vanderveer, 1980), and radiated underwater sounds such as propeller cavitation (Howard & Ballas, 1983). These results have shown that listeners are surprisingly accurate in identifying the sound source or in characterizing some specified attribute of the sound source. A number of recent investigators have suggested that this capability may prove useful for developing automatic classification (Gorman & Sawarti, 1985) strategies. For example, in one case, human experts were used to develop an intelligent, knowledge-based system for passive underwater surveillance (Nii & Feigenbaum, 1982), and in another, human listeners were used to identify a set of acoustic features for classifying active sonar returns (Gorman & Sawarti, 1985).

Neural networks have also been demonstrated to perform a wide range of classification and pattern recognition tasks. However, neural networks may perform better when integrated with human classification capabilities or vise versa. To demonstrate this, ARD has investigated how humans and networks interact by developing and testing a prototype system in which people and networks act jointly and individually to classify signals. The objectives are to measure the relative performance of humans, networks and the combination of the two, to find out what tools are desirable for the operator to use to make classification decisions, and to determine what kind of decisions the operator is willing to let the network make.

To accomplish these goals, ARD developed a neural network based prototype system using signals from a previous contract with the Naval Air Systems Command. The signals set used in the experiments conducted as part of this project were selected from a group of 15,744 signals collected in a laboratory at the Naval Coastal Systems Center in Panama City, Florida. The signals were divided into two groups representing a very clean set and a set where the SNR was reduced to 8.5dB. Test subjects were asked to identify three characteristics of the signals using four basic tools and a neural network classification system. System utilization and classification performance was automatically recorded during each session for post-experiment analysis.

Three experiments were conducted. In the first experiment, test subjects were asked to identify the thirty-six signals using four traditional tools, but not the networks. Subjects were automatically presented with the time domain waveform of the signal and allowed to call up the frequency domain plot or spectrogram of the signal. A time windowing function was also provided to allow the user to zoom in and take a closer look at specific portions of the time domain signal. In addition, the subjects could listen to the sound as many times as they desired. Experiment two recorded the performance of the neural networks operating alone, without help or intervention of the operator. In the third experiment, test subjects were allowed to use the classification abilities of the networks to aid in the decision-making process in addition to the tools used in experiment one. The experiments are discussed in detail in Section 5 and the analysis of the experiments is discussed in Section 6.

1.3 Artificial Neural Networks

Traditional general-purpose digital computers have a fundamentally serial architecture. This architecture, sometimes known as a von Neumann architecture, is characterized by a single, very powerful processor which executes a set of instructions sequentially in a step-wise fashion. Dramatic advances in the speed of these machines have been achieved primarily through large-scale integration which effectively increases the density of system components. There is a growing awareness in computer engineering, however, that current technologies are approaching an upper bound on processor and memory speed; and, that further improvements in system throughput must be achieved by adding processors rather than by increasing the speed of individual processors. These developments have led to the recent burst of research activity on parallel architectures. Artificial neural networks (ANN) or connectionist networks reflect one approach to massively parallel architectures of this sort.

Many ANNs have been designed to imitate some of the very gross properties of living nervous systems. Hence, they are characterized in terms of a set of very simple, neuron-like computational elements which are massively interconnected to form a network capable of performing complex computations. Computations in such a network are carried out in parallel with each unit operating concurrently with the others. The output of each element or processor is typically a non-linear transform of the weighted sum of its inputs (either from other network elements or from measurements external to the network). Hence, the actual computation carried out by the network is determined by the weight values for the interconnections between units and the non-linear function. The design of these systems not only achieves substantial improvements in processing speed over conventional systems, but also leads to a number of other useful characteristics as well. Among these is a self-organizing capability which permits them to learn to solve a particular problem. During training, the network is presented with a series of signals, each paired with a desired output or target value. Various learning algorithms exist which specify how the network weights are adjusted to minimize the overall error between the computed and target output. Once a network has learned a mapping, it may be used for direct classification or for feature extraction using unfamiliar signals. This characteristic obviates the need to specify signal features a priori.

For this project, ANNs have been exploited as a near perfect classifier for all the clean and some noisy signals. ARD purposefully degraded the signal set to the point where the networks would not be a perfect classifier to encourage the users not to use the networks exclusively. Section 2 of this report describes the signal set and Section 4.0 describes the neural networks in detail. It should be noted, however, that the neural networks can be trained to classify the clean signals with 100% accuracy. As noise is added to the signals, the network's performance drops off very slowly. Even after the SNR has been reduced to 8.5 dB, performance does not drop off precipitously as one might expect. In fact, network performance never fell below chance even when the signal-to-noise ratio was reduced to -4 dB. Section 5.3 contains a set of illustrations which better depict the significance of this finding. For this reason, ARD believes that neural networks may be very effective at classifying signals from real-world environments.

2.0 SIGNALS

2.1 Signal Parameters

The goal of this research was to evaluate the interaction between a human operator and an acoustic classification system containing several tools to aid in identifying acoustic signals. Since the interaction itself was of highest interest, a controlled data set which would not complicate the evaluation process was desired. To this end, the signal set employed in the experiments was part of an extensive set of signals collected in a laboratory setting at the Naval Coastal Systems Center (NCSC) in Panama City, Florida. Although the signals were collected under laboratory conditions, they represent significant and realistic parameters in the realm of underwater acoustics.

The signals were sonar returns from the insonification of two steel targets which are scaled models of mines. Each target had a unique shell thickness to diameter ratio. One shell was five percent of the outside diameter of the target, and the other was ten percent. The targets were constructed to within 0.005 inches of the original specification. Detailed drawings for the specification are provided in Figure 2-1A. Figure 2-1B is a photograph made at the time the targets were inspected for tolerances. Figure 2-1C is a photograph of the targets after data collection was completed.

The two shell thicknesses were used in combination with different interior contents and angles of insonification to give the signal set realistic attributes. The three interior contents were air, water and a solid epoxy. The angles of incidence were 90 degrees (the target suspended broadside to the transducer and hydrophone), 45 degrees and 0 degrees (end on). Varying these three parameters produced a set of 18 signal classes:

2 Shell Thicknesses	x 3 Angles	x 3 Contents	= 18 classes
Five Percent	90 Degrees	Air	
Ten Percent	45 Degrees	Water	
	0 Degrees	Solid	

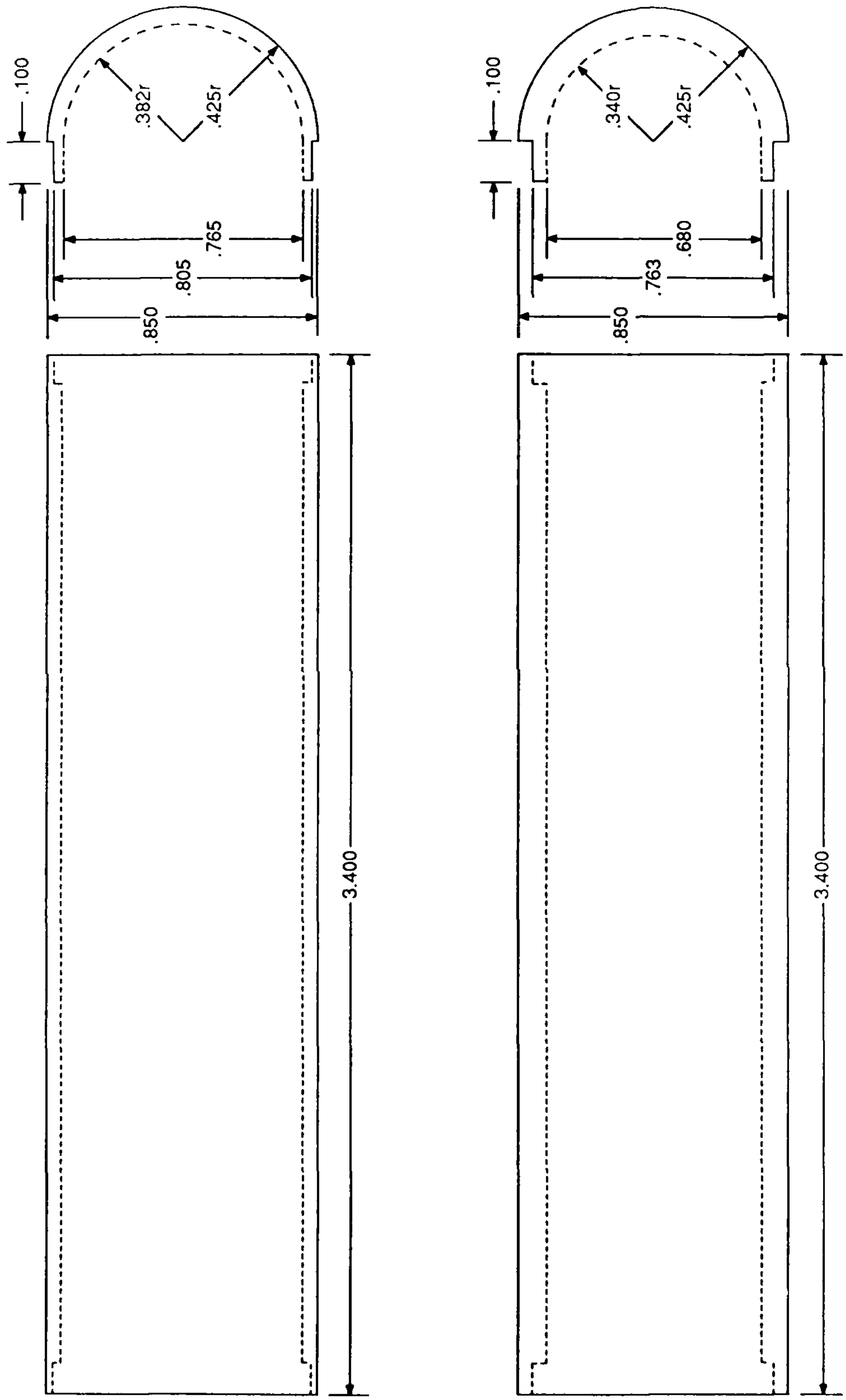


Figure 2-1A Detailed Target Drawing

ARD

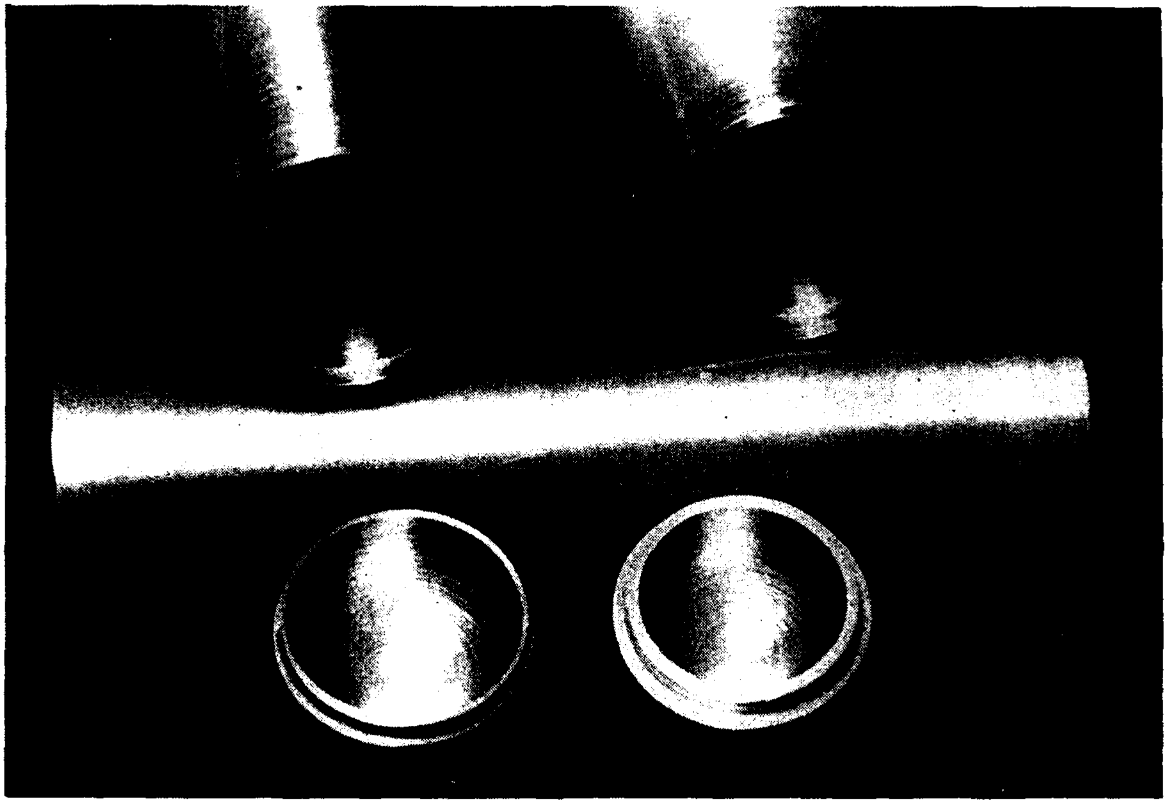


Figure 2-1B Targets Before Final Assembly

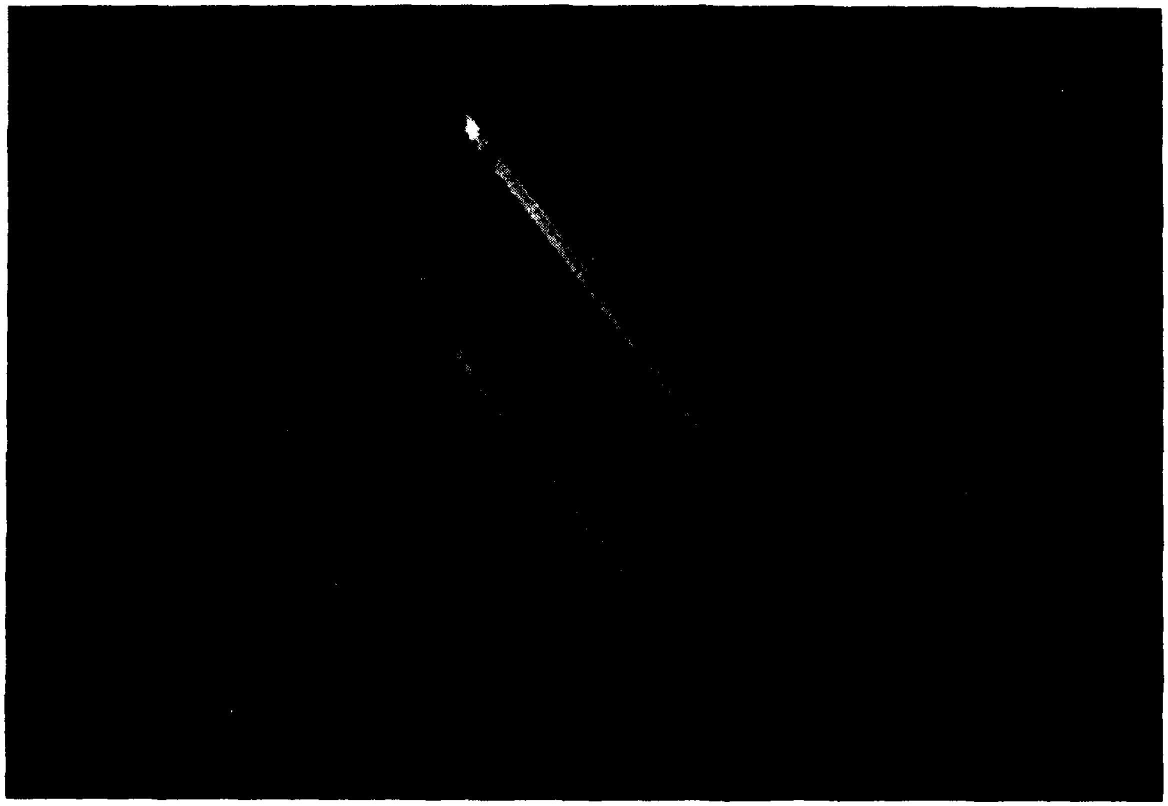


Figure 2-1C Targets After Data Collection

At each of these conditions, 32 signals were collected to allow a set of averaged signals to be constructed to produce very clean signals. This set of parameter combinations was sufficient for the experiments conducted for this project, but it represents only a portion of the complete set collected. For the sake of brevity, only the information relevant to the parameters used in the current work will be described.

2.2 Data Collection

The signal set was collected in a facility at NCSC. To perform the actual collection, each target was suspended in a 10' x 10' x 7' tank and insonified with 6 cycles of a 200 kHz sinusoid. The tank is shown in Figure 2-2, and the collection hardware is shown in Figure 2-3. The reflected acoustic signals were sampled at 2 MHz and digitized over 12 bits, resulting in an amplitude resolution of 4096 discrete values. One thousand and twenty four (1024) samples were obtained for each signal.

2.3 Signal Conversion

The raw signals were converted to produce signals in the format needed for human experimentation and neural network training. Since the signals were digitized at 12 bits, using 11 bits for amplitude and one for sign, the first step of the conversion resulted in signals of ASCII data in the range (-2048, 2047) as shown in Figure 2-4A. Any DC component (offset from zero), was removed by subtracting the mean of each signal from all points in that signal. This made the mean of every signal zero. The next step in the process was to normalize the signals by adjusting the amplitudes in each signal to a range of (1,-1). This step was taken to equalize the amplitudes of all the signals in the set. This was necessary to preclude the subjects from using differences in the amplitude of the signal as a cue to any of the parameters. This method of equalization was just one of several possible solutions. It was chosen as the simplest method likely to accomplish the objective. To make this adjustment the maximum absolute value of the points in each signal was determined. The maximum absolute value varied considerably from class to class, and very slightly from signal to signal within a class. All points in the signal were divided by this absolute value, making the range of amplitude values (1,-1) and

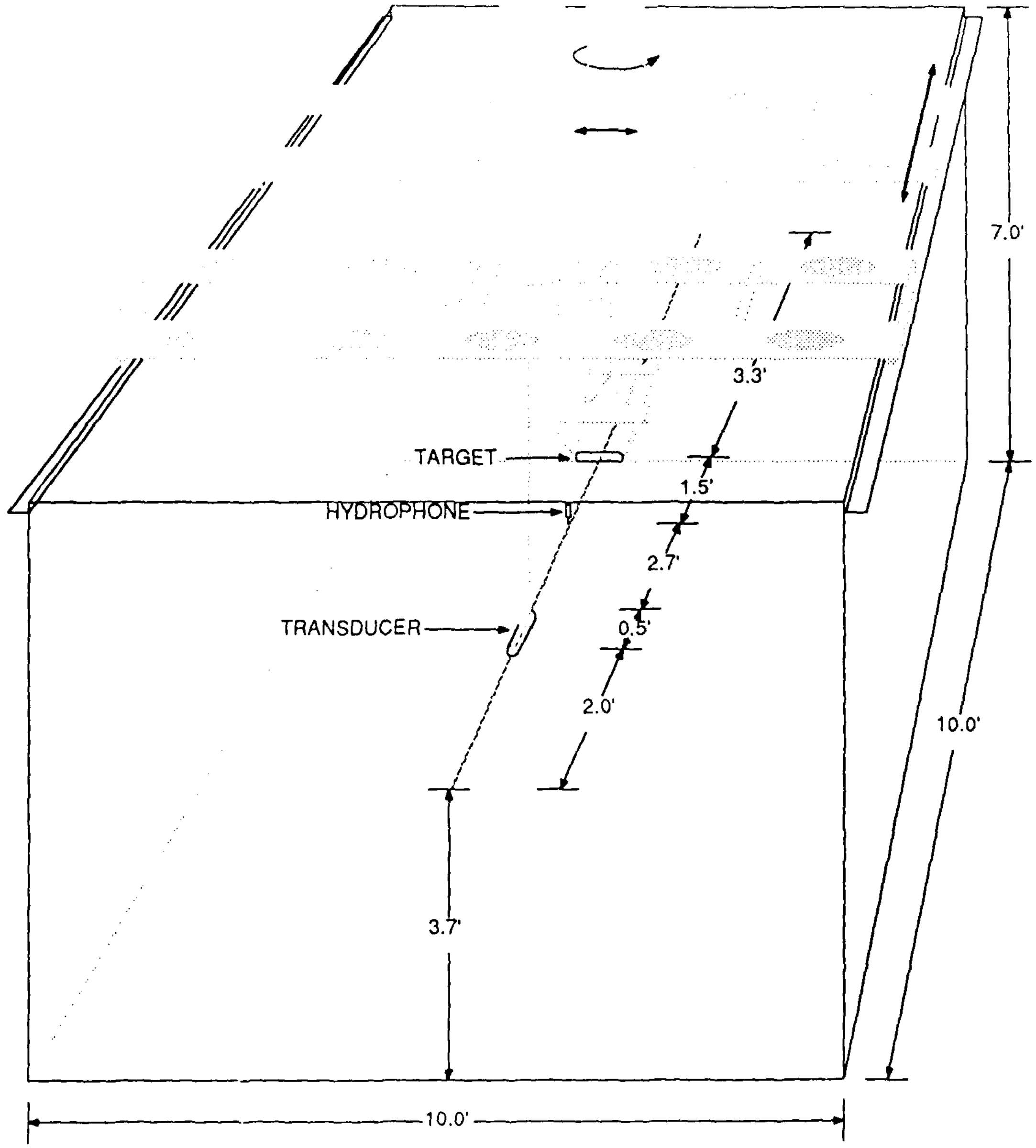


Figure 2-2 NCSC Tank for Acoustic Data Collection

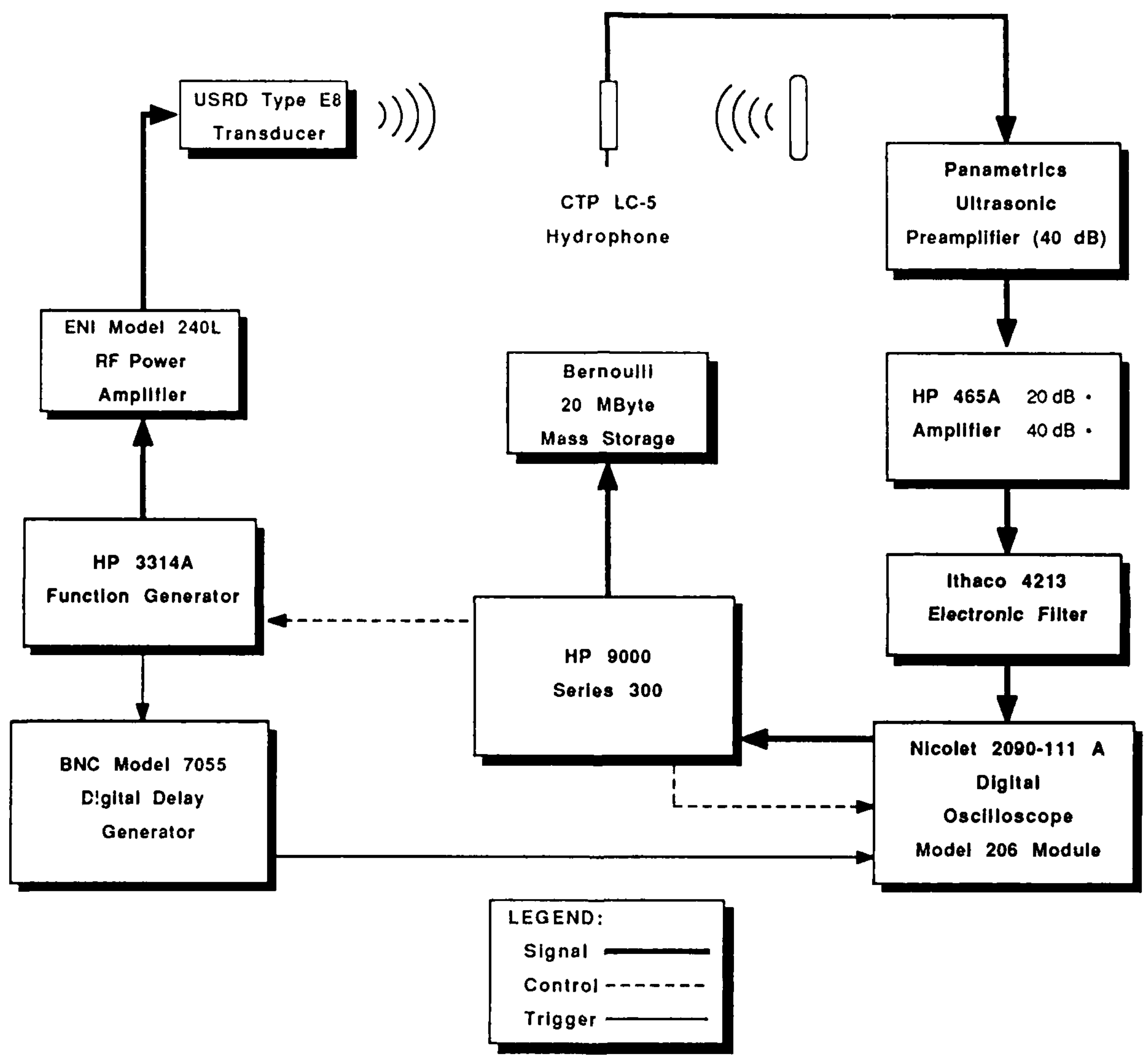


Figure 2-3 NCSC Data Collection System

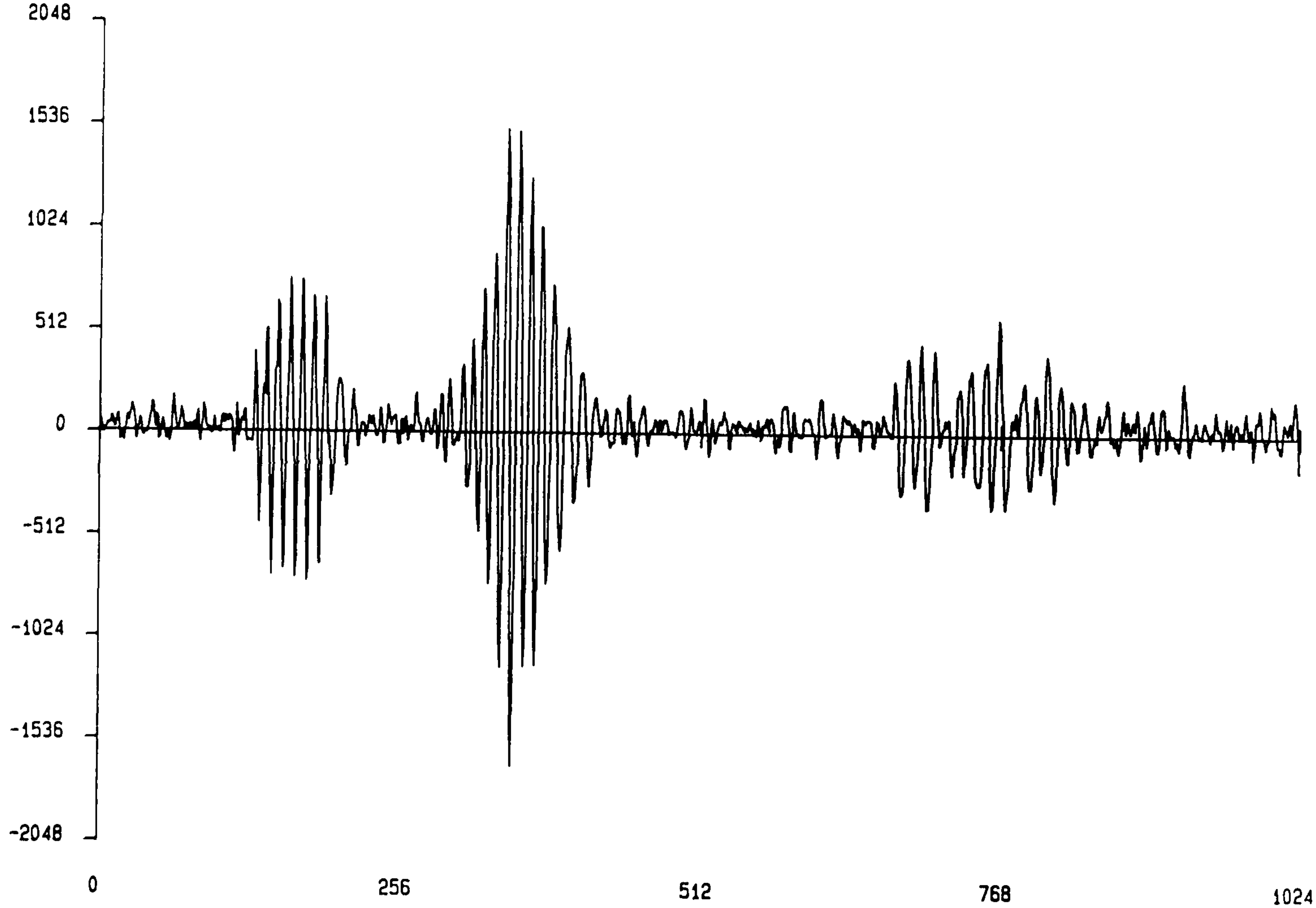


Figure 2-4A One Instance of S145
in Original ASCII Form

guaranteeing that at least one point in each signal was either 1 or -1. An example of a normalized, mean-zero adjusted signal is shown in Figure 2-4B.

Although the signals were "normalized" to the range (1,-1), they were still 1024 points long. Due to the rotation of the targets in the tank, and to small differences in position each time a target was suspended in the tank, the initial specular return (reflection of the six-cycle sinusoid) did not occur at the same time in each class of signal. In addition, late in each signal, a reflection from the surface of the water appeared. This was due to the geometry of the tank, as shown in Figure 2-5. Both the location of the surface return and its amplitude were related to the class of signal. Therefore, the surface return had to be eliminated from the signals to preclude its use as a cue to the class of signal. "Standardization" of the signals was the process of aligning each class of signal at its specular and eliminating the points which included the surface return.

2.4 Signal Standardization

Standardization was a four-step operation. First, the signals were time synchronized (aligned) relative to their initial specular return. Second, the surface return was removed by deleting points from a predetermined location to the end of the signal. Third, only for the averaged signals played audibly to the subjects, the signals were ramped up near the specular and down before the surface return. And fourth, leading zeros replaced the noise at the beginning of the signals, and padded the end of the signals to 500 points.

In order to align the signals, the specular had to be precisely located in a small level of noise. The automatic method used to find each signal's specular was a mean window algorithm. The algorithm consisted of taking the mean of the absolute values of the first fifty points in a signal, which were known to be noise, multiplying the mean by a gain factor and comparing the product to the absolute value of each point, starting with the second point. If a point was larger than the product, the next three points were checked. If three of the four points were larger than the product, then the first point which satisfied the criterion was marked as the first point in the specular. Only three points were required to meet the criterion to allow for one of the points to be close

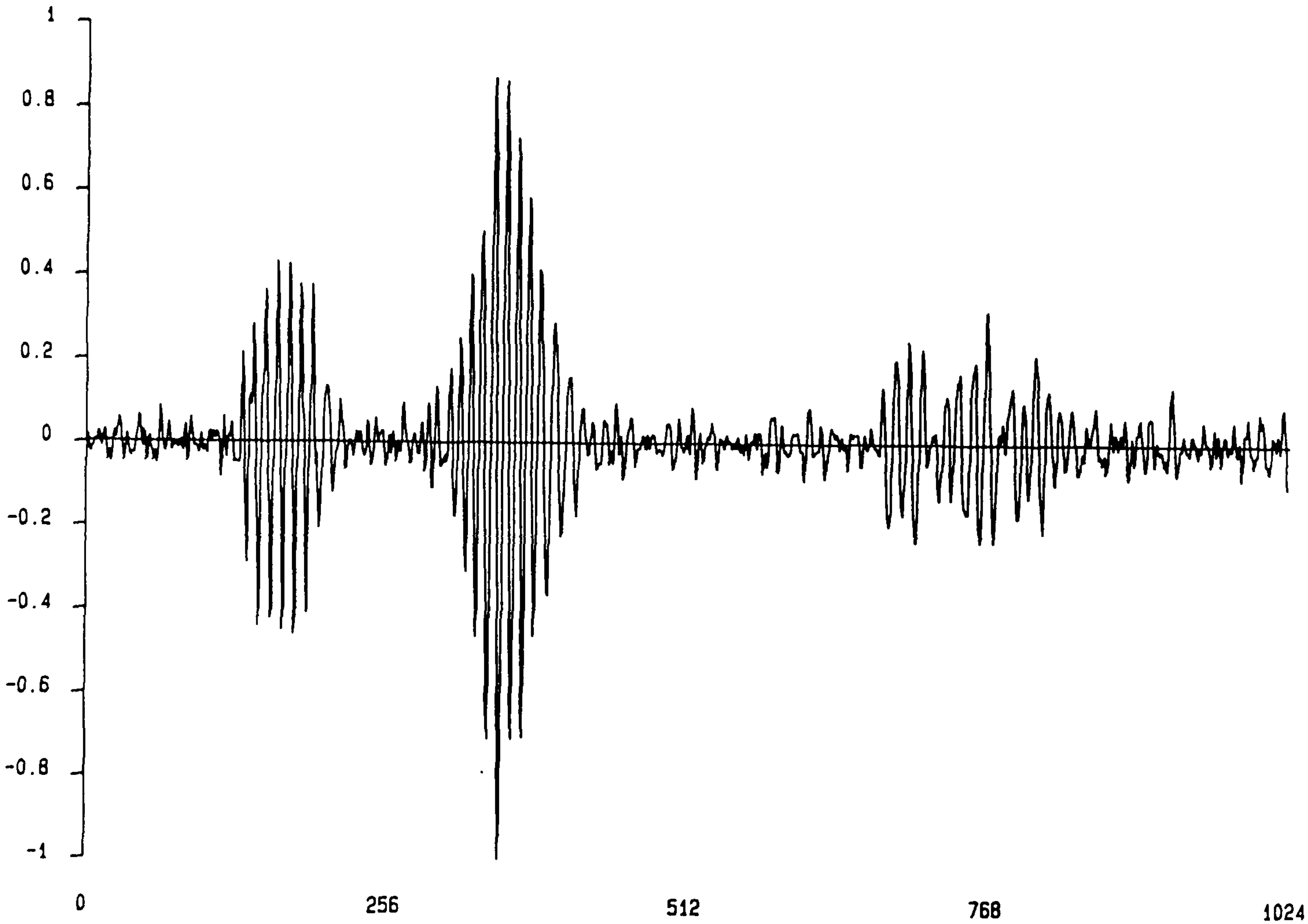


Figure 2-4B One Instance of S145 After It Was Adjusted to Mean = 0 and Normalized to (-1,1)

ARD

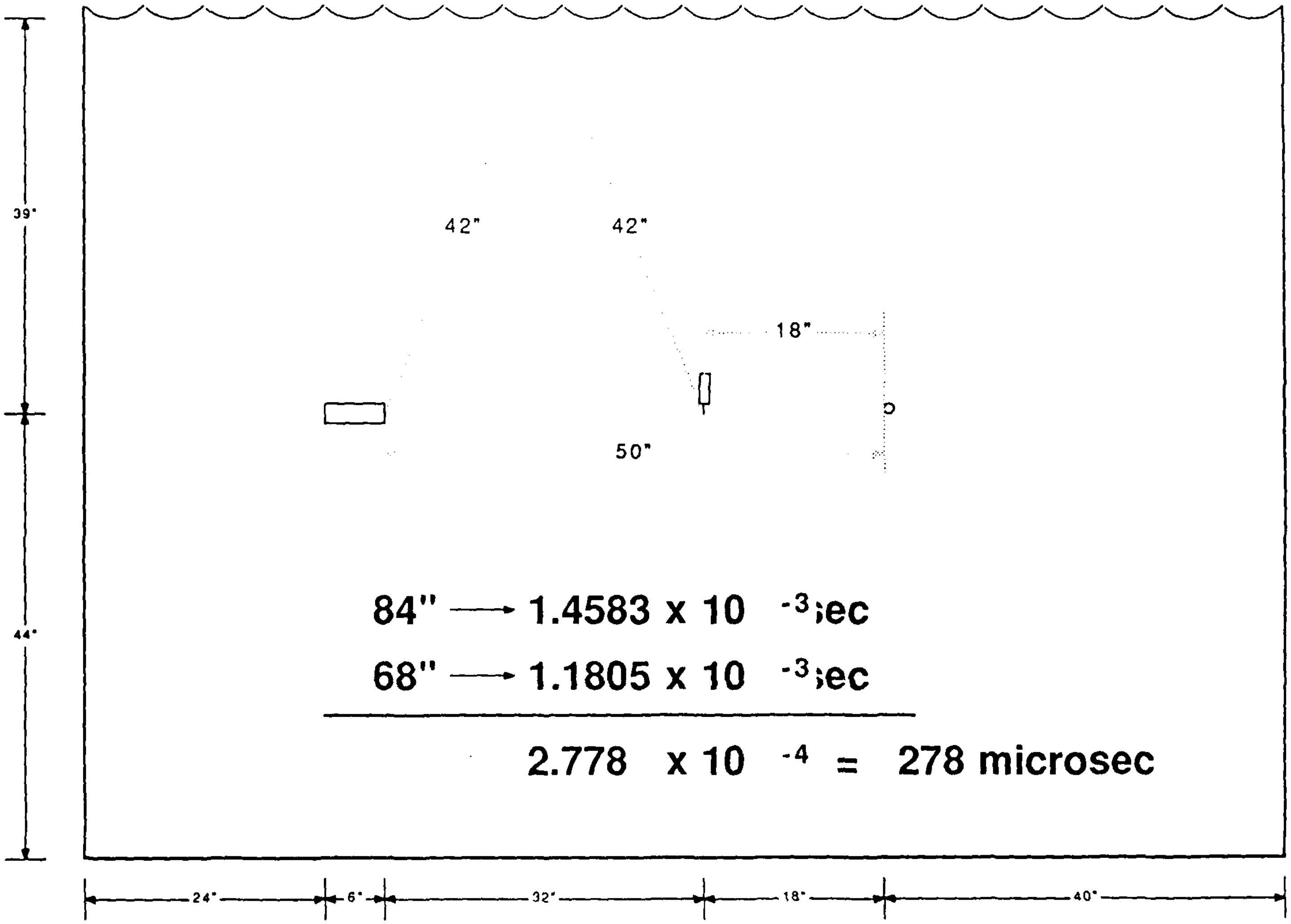


Figure 2-5 The Surface Reflection

to zero as the signal crosses the x axis. Four points were checked because random noise could sometimes exceed the product of the mean and the gain factor.

The surface return in each class of signal was found using a combination of visual inspection and the geometry of the tank. A fixed number of points between the specular and the surface return were calculated for each class. The minimum number of points between the specular and the surface return was applied to all signals. All points more than the minimum number beyond the specular were eliminated. The entire signal was then shifted to the left by dropping leading points until the specular began exactly 25 points into the signal.

Finally ramping was applied, but only on a separate set of averaged signals which were played audibly for the subjects in the experiments. The ramping started five points before the specular and continued through nine points after the first point of the specular, giving a ramp of fifteen points. The purpose of the ramp was to gradually introduce the main energy of the signal. This prevented spurious aliasing caused by the sudden onset of a high level of energy. The actual ramp was performed by multiplying each of the fifteen points by a linearly increasing factor between zero and one. In this way the points at the beginning of the ramp were multiplied by a smaller factor than those at the end, thus giving the required graduation of energy. Conversely, the end of the signal was linearly down ramped, starting at the fifteenth point before the end of the signal. The down ramping was done to smoothly taper the energy level down to zero. Between the end of the ramped points in the specular and the down-ramped points at the end of the signal, the points were simply copied from the original version of the signal to the time synchronized signal. After the ramped points at the end of the signal, zeros were used to pad each signal out to 500 points.

For the purposes of the three experiments, the 32 instances of each normalized, standardized signal were split into groups of eight signals. The groups were averaged into four signals: two to be used for training and two for testing. The averaging was accomplished by averaging each of the 500 points in the signals across the signals. The ith point in the resulting averaged signal was the result of summing the ith point of each signal and dividing the sum by 8. Averaging the signals produced a cleaner example, and a higher signal-to-noise ratio than the original instances of the signals. An example of an averaged, "standardized," but unramped signal is shown in Figure 2-6.

A comparison of Figure 2-4A and 2-6 best illustrates the effect of the signal conversion process applied to the signals. The 18 classes of signals are shown in averaged (over eight instances), normalized, standardized, unramped form in Appendix A.

To facilitate references to the signals a coding convention was adopted. Signals are referred to by up to six characters. The first character is either A, S, or W, identifying the content as air, solid, or water. The second character is either 5 or 1, identifying 5% (thin) or 10% (thick) shell thickness. The third and fourth characters identify the angle (e.g., 45). The fifth and sixth characters are 20 which is a shortened version of the 200 kHz frequency of insonification. At times the fifth and sixth characters are not present. For space reasons in some charts the angle is identified with a single digit as 9 (90 degrees), 4 (45 degrees), or 0 (0 degrees).

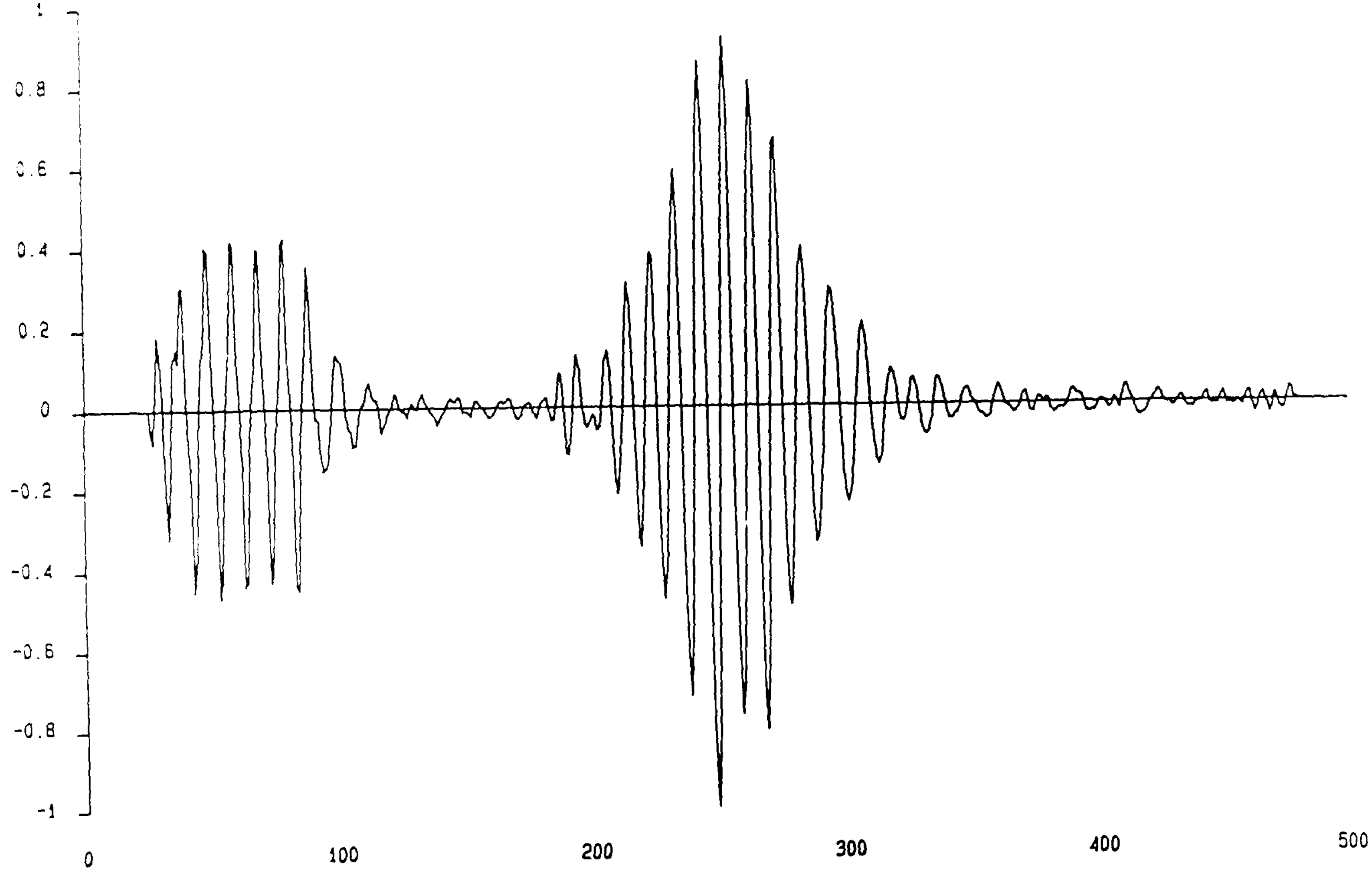


Figure 2-6 8 Instances of S145 Averaged Together After Finding Specular, Shifting, and Cutting Off Surface Return

3.0 SOFTWARE

3.1 General Description

Software to implement the experiments was written largely in the C programming language on a Micro Express 386/25 with 640k of RAM and an 80Mb hard disk drive. The software uses a largely graphical interface, and was implemented on the EGA standard. The specific hardware configuration and the way the subjects used it is shown in Figure 3-1. As can be seen, this is a two monitor system with all the graphics on one VGA monitor with 16 colors and 640 X 480 resolution and all the textual information is displayed on a monochrome monitor. A Data Translation 2801A D/A board was used to convert the digital waveforms to analog signals and play at 10,000Hz. A low-pass anti-aliasing filter with a cutoff frequency of 5,000 Hz was used to eliminate high frequency artifacts. An NAD 7225PE receiver amplified the signals, and the subjects heard them on Sony MDR-V6 headphones.

The software developed for this project was used to conduct a series of experiments and is not a required deliverable. Should ONR want a copy, it could easily be made available. Care was taken to ensure the software was of very high quality. It would also be possible to reuse this software on other projects or to modify it for use in a Phase II follow-on to this effort. The only additional work necessary to turn this into a deliverable would be to develop a users guide and installation instructions.

3.2 Software Development

Software development was divided into four phases: 1) development of the digital signal processing tools and graphics interface, 2) development of the menu interface and instructions, 3) development of the neural networks and 4) the development of the experimental software to collect and analyze the data. A fifth aspect of software development was the preparation of the signal set which is discussed in detail in Section 2. The software was written largely in Microsoft C Version 5.1 under DOS Version 4.01. The graphics were developed using EGA graphics routines from Connell Graphics Version 3.0. The menu system

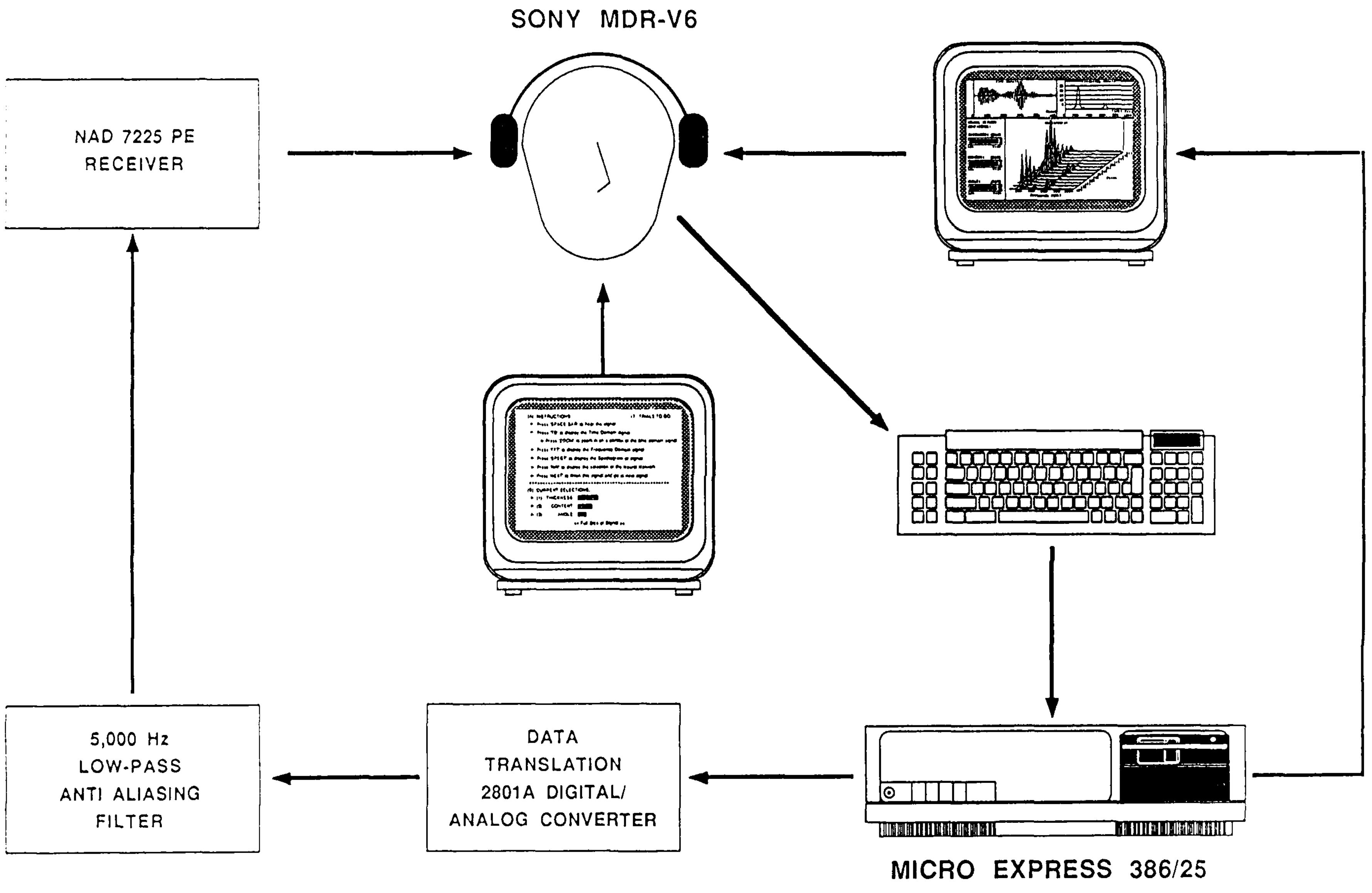


Figure 3-1 Experimental Hardware Configuration

ARD

was developed using a Hercules monochrome video system. The FFT algorithm was adapted from the book Digital Spectral Analysis with Applications (Marple) and written in Microsoft Fortran Version 4.1. Data Translation's PCLAB library Version 3.01 was used to do the A/D conversions. The system was designed and implemented with the help of two human factors engineers to ensure the highest level of user acceptance and usability. This approach was highly successful in that there were virtually no questions by the users on the intent or function of the system.

3.2.1 Digital Signal Processing (DSP) Tools

Originally, ARD had planned on using a set of tools called the Interactive Laboratory System from Signal Technologies Incorporated (STI) to handle the DSP portions of the software. As advertised, the software from STI should have been able the integrate with user-developed software on a PC. However, in actual practice, this was not possible for a 386 class PC. Given that it was imperative to have a core set of DSP tools imbedded in the experimental software, ARD developed its own system that allows a user to display a time domain plot, a frequency domain plot and a spectrogram of the signals used in the experiments.

The time domain plot, as shown in Figures 3-2 and 3-3 (clean and noisy versions of a signal), was automatically displayed each time a new signal was brought into the system. This is the 500-point representation of the signal after going through the manipulations described in Section 2. The decision to make the time domain appear automatically was based on the notion that in Experiment 3 the users may be tempted to only use the data provided by the neural networks. Since the purpose of the study is to analyze how users interact with a system employing a neural network classifier, ARD decided that some type of induced interaction might be necessary in such a circumstance. To keep the experiments as similar as possible, the time domain signal was automatically displayed in Experiment 1 as well.

In addition to the standard display of the time domain plot of the signal, users were allowed to zoom in on any specific portion of the signal to gain a higher degree of resolution for that portion of the signal. This was done to

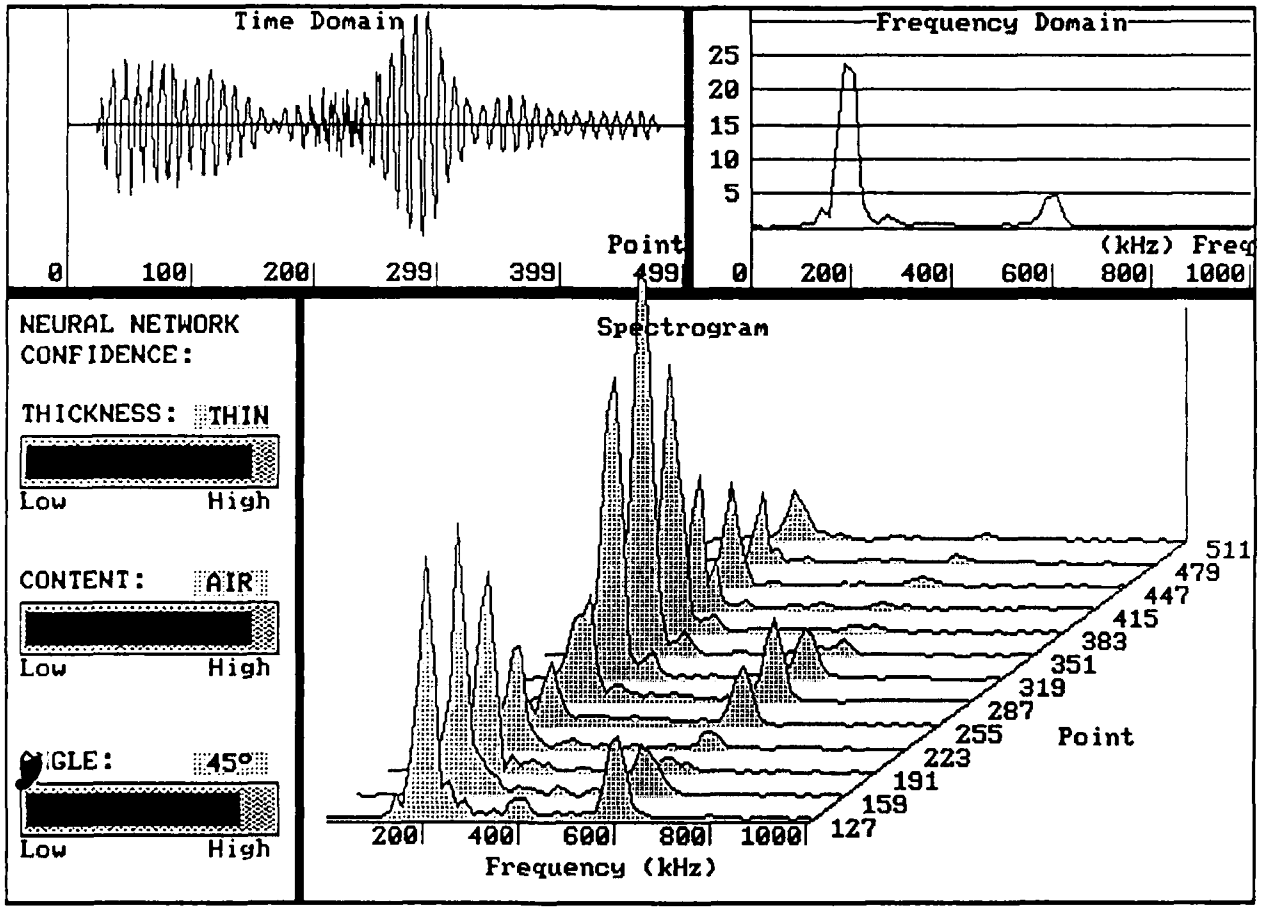


Figure 3-2 Graphics Display From Experiment 3 Showing a Clean Signal

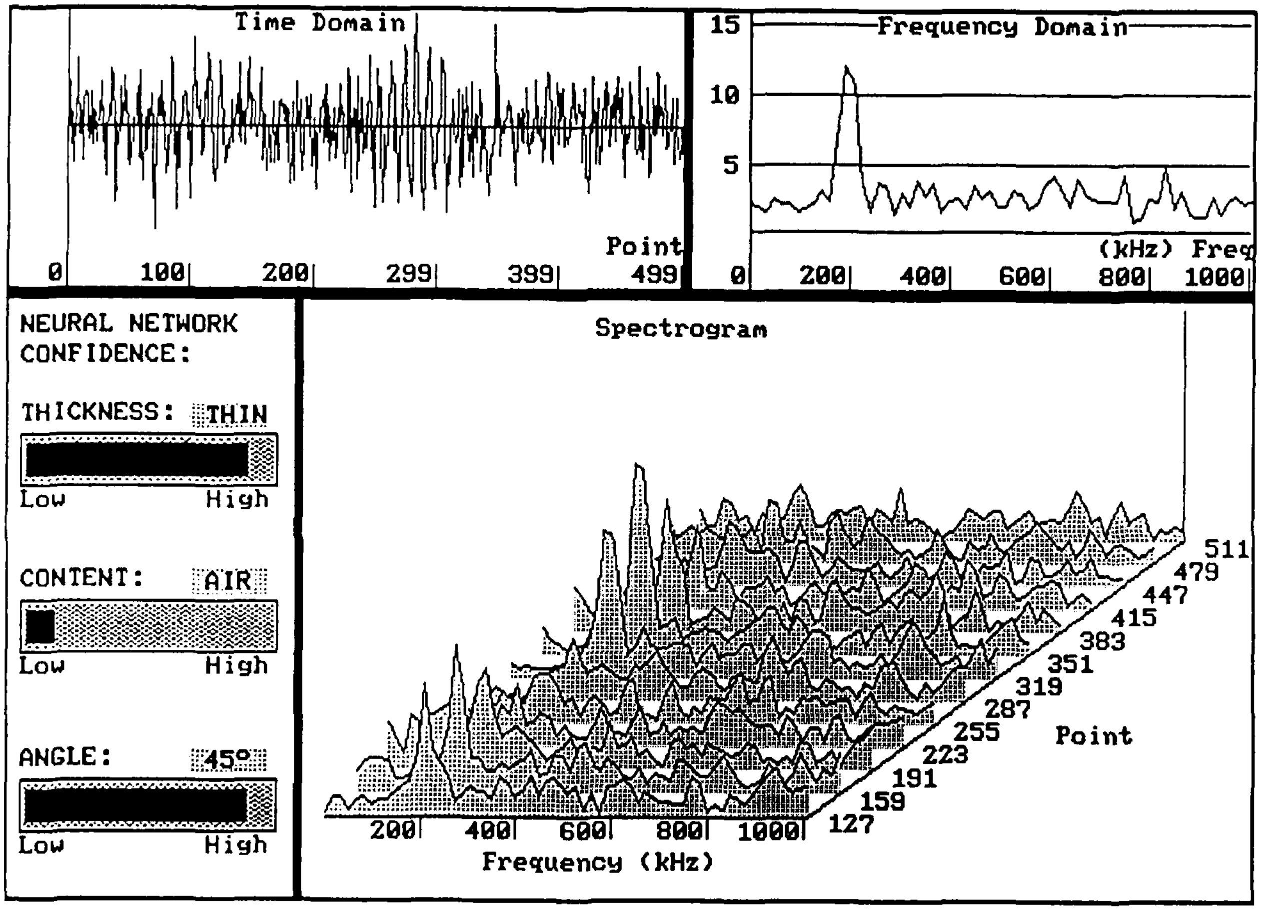


Figure 3-3 Graphics Display From Experiment 3 Showing a Noisy Signal

ARD

allow users to try and determine for themselves if any specific portion of the signal held the key to its identity. Each time the time domain signal is redisplayed, the frequency domain plot is also regenerated using the pared down data. The spectrogram display was not affected by the "ZOOM" function.

The frequency domain plot was generated by taking the 500 point time domain signal and computing the Fast Fourier Transform (FFT). This produces 256 complex values, and we took the absolute value of each complex number to get magnitude values. We then averaged each four points to bring the resolution of the frequency domain plot down from 256 points to 64 points to match the resolution of the spectrograms. The resulting frequency plot was then displayed in a window on the graphics monitor. This tool was only displayed on demand by the user.

The spectrogram is created by dividing the time domain signal into 13 overlapping windows of 128 points each and computing an FFT on each window. the results are then displayed in an overlapped fashion in a window on the graphics screen. This tool was only displayed on demand by the user and will be discussed in Section 5., Analysis.

3.2.2 Menu Interface

The menu and instruction screen was separated from the graphic displays to keep from overcrowding a single display and to keep from switching back and forth between instructions and the tools. ARD's human factors engineers felt this was the only way to ergonomically handle the amount of data without developing the system in a windowing environment. As a result, all the system functions were clearly described on line for the user and space was provided for the users to enter their responses to the classification questions posed by the system. This is illustrated in Figure 3-4A.

Primarily, this screen told the users what to do at each step of the process. It also told the users how many more signals would be analyzed before the session ended. It allowed the users to record their selections for the three key parameters to be identified from the signal. Keys were labeled so that the

(A) INSTRUCTIONS: 17 TRIALS TO GO

- Press PLAY to hear the signal
- Press TIME to display the Time Domain signal
  - Press ZOOM to zoom in on a portion of the time domain signal
- Press FREQ to display the Frequency Domain signal
- Press SPEC to display the Spectogram of signal
- Press NET to display the selection of the Neural Network
- Press NEXT to finish this signal and go to next signal

>>═══════════════════<<

(B) CURRENT SELECTIONS:

(1) THICKNESS:

(2) CONTENT:

(3) ANGLE:

<< Full Size of Signal >>

Figure 3-4A Menu Interface Before the User Attempts to Classify the Signal

(A) INSTRUCTIONS: 17 TRIALS TO GO

- Press PLAY to hear the signal
- Press TIME to display the Time Domain signal
  - Press ZOOM to zoom in on a portion of the time domain signal
- Press FREQ to display the Frequency Domain signal
- Press SPEC to display the Spectogram of signal
- Press NET to display the selection of the Neural Network
- Press NEXT to finish this signal and go to next signal

>>═══════════════════<<

(B) CURRENT SELECTIONS:		(C) ACTUAL SIGNAL:	
(1) THICKNESS:	THICK	(1) THICKNESS:	THIN
(2) CONTENT:	AIR	(2) CONTENT:	AIR
(3) ANGLE:	0°	(3) ANGLE:	0°

Press NEXT to Continue . . .

Figure 3-4B Menu Interface After the User Attempts to Classify the Signal first five function keys would allow the user to: F1, redisplay the time domain signal after issuing a zooming command; F2, zoom in on a specific portion of the time domain signal; F3, display the frequency domain plot of the signal, F4, display the spectrogram of the signal; and F5, display the neural network analysis/classification of the signal (in Experiment 3 only). The space bar was labeled "Play" to indicate that pressing this key would audibly play the signal.

The bottom portion of the screen was reserved for the users to record their responses and for the users to get feedback on the correct classifications Users could select any of the categories while simultaneously using the tools described above, until the "Next" key was pressed, as shown in Figure 3-4B. Once the "Next" key is pressed, the graphics display was frozen and the correct classifications were presented to the user. This feedback was not only important in helping the user to learn the signal set early in Experiment 1, it continued to help to improve the user's performance throughout the experiments.

3.2.3 Neural Network Software

Networks were trained using the backpropagation paradigm. [A short description of Backpropagation (BPN) is contained in Section 4]. The networks were trained using the 500 amplitude points of the time domain signals as input to the networks. The number of hidden nodes was fixed at eight after several training runs to determine the optimum number. There were eight output nodes to account for the eight principal parameters being classified in the experiments, two for shell thickness, three for interior contents and three for angle of orientation. Training was carried out on a Compaq 386/20 using an HNC accelerator board. The resulting weights for the trained networks were transferred to the Micro Express 386/25 used for the experiments along with an ARD developed BPN to carry out operational runs of the network.

During Experiment 3, the network was run for each new signal if the user requested it. This run was conducted in real time and the results displayed in the form of a bar graph (Figure 3-2 and Figure 3-3) to indicate the relative confidence of the network that it had developed the correct response. This confidence level was related to the strengths of the activations on the output nodes. For example, the network was trained to learn the desired output of 0.99 for node one for a thick-shelled target. If the actual output was 0.90, then the bar graph would show a very high confidence. However, if the network produced an actual output of 0.60 then the confidence would be significantly lower. In actual practice, the confidence was quite low on several of the noisy signals even when the networks were correctly identifying the objects. This was intended to reduce the test subjects dependence on the network's classification analysis.

3.2.4 Experimental Control Software

Several small applications were developed as necessary to support the actual conduct of the data collection effort. The first step was to make it easy for the user to access the system and record the results. Two log-on screens were developed to clearly identify Experiments 1 and 3. Batch jobs were run at the end of each session to back up the data. During each session, the sequence of key strokes was recorded for post experiment analysis. Once all the subjects were run through both Experiment 1 and 3 additional software was written to break down the data in various ways to prepare it for statistical analysis.

4.0 NEURAL NETWORKS

4.1 Introduction

In addition to the two human experiments, ARD carried out an analysis based solely on the performance of artificial neural networks (ANNs) using both clean and noisy signals. As mentioned in the introduction, ANNs can be useful for classifying data with minimal prior knowledge regarding specific features of the data. We have applied this technology to the problem at hand to gain a better insight into the interaction of humans and neural network based systems.

4.2 Network Training by Backpropagation

The networks used were trained for sonar classification using the method of Rumelhart, Hinton, and Williams (1986). This method, called the generalized delta rule, enables the inter-unit connection weights to be adjusted empirically on the basis of training experience and is the basis for the backpropagation paradigm. During training, pairs of input and target or desired output vectors are presented to the networks. For each pair, a set of output values is computed and an error signal is determined for each output unit which is based on the difference between the observed and target values. This is shown schematically for a three-layer network in Figure 4-1. Weights

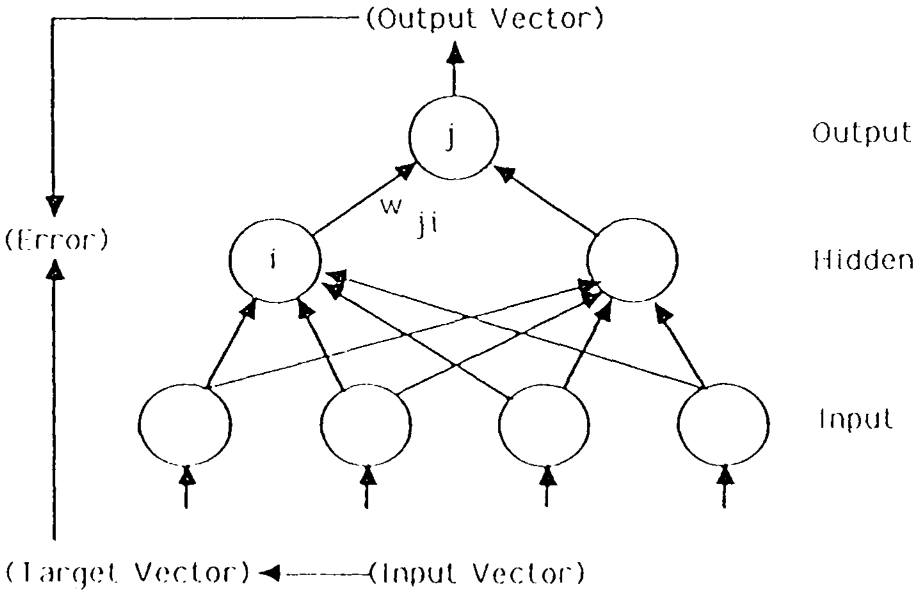


Figure 4-1 Error Back Propagation

between each output unit and the hidden units are then adjusted by an amount proportional to three quantities: 1) the error for that output unit, 2) the output of the hidden unit, and 3) a learning rate parameter (between 0.0 and 1.0). The learning rate parameter serves to avoid overcorrection thereby preventing oscillations in the weights as the outputs converge to the target values. To illustrate this process, consider the weight between output unit j and hidden unit i, w_{ji}, shown in Figure 4-1. The adjustment term for this weight, Δw_{ji}, is simply the product of the learning rate, η , the error signal, δ_j, and the output of unit i, O_i

$$\Delta w_{ji} = \eta \delta_i O_i$$

where the error signal is the difference between the target value and the actual output, weighted by the derivative of the nonlinearity used to "squash" the output. For the logistic squashing function used in this research, the error term for an output unit is given by

$$\delta_j = (t_j - O_j) O_j (1 - O_j)$$

A similar adjustment must be applied to the weights between the input and hidden units. Unlike the output units, however, target values cannot be specified directly for the internal or hidden units. To estimate the error term for each hidden unit we apportion the observable or output error among the hidden units in proportion to the weights between the hidden and output units. This estimated error is again weighted by the derivative of the squashing function and for hidden unit k, is given by

$$\delta_k = O_k (1 - O_k) \sum_j w_{jk} \delta_j$$

where the sum $\sum w_{jk} \delta_j$, is taken over the j output units which connect to this unit. It has been shown theoretically that the generalized delta rule serves to minimize the sum of squared errors between the observed and target

ARD

signals by gradient descent in the weight space. Similar adjustments are made in the bias or threshold terms for each unit. Repeated application of this process produces a trained network which maps the input data set to the target data set. This self-learning capability makes backpropagation well suited for acoustic classification problems in which the functional relationship between the input/output mapping is not understood analytically.

5.0 EXPERIMENTS

5.1 Introduction

A two-monitor PC configuration was used for the human experiments. This allowed the display of all the graphical representations of the signals on one monitor while all textual materials were managed on the second. The users had four base tools and the neural network classification system (Experiment 3 only) to use in making their personal judgements on the parameters to be identified from the signals. The four base tools included a time domain plot of the signal (which included a windowing feature to allow the user to select a portion of the signal for display or listening), a frequency domain plot of the signal, a spectrogram plot of the signal, and of course, the ability to hear the signal (or portion of the signal) as many times as the users wanted.

Each subject ran ten sessions of Experiment 1 and five sessions of Experiment 3. During the course of each session, subjects were presented with 32 signals, one at a time. As described in Section 2, 18 signals were clean and 18 signals were noisy (SNR reduced to 8.5 dB). As each signal entered the queue, the subjects had the option to invoke any of the available tools or to select any of the three parameters being classified. To enter a selection, the user pressed one of the specially labeled keys on the numeric key pad. The selection was registered at the bottom of the monochrome display beside the appropriate label as shown in Figure 3-4B. Once all three selections were entered, the subject pressed the "NEXT" key to check his answers.

Instructions displayed on the text monitor controlled the information displayed on the graphics monitor. Responses and commands given by the user were entered via a standard keyboard with predefined keys, clearly labeled as to their meaning and intended function. The following discussion describes how the experiments were conducted. Refer to Figures 3-2 and 3-4 as an example of the information the system displayed for the user during each session.

5.2 Experiment 1: Operators Using Base Tools Without Neural Networks

When the test subject sat down to begin an experiment, the computer was off. Turning on the power strip with all the system components plugged in turned everything on. A small batch file automatically ran to change to the correct subdirectory on the system where the experimental software was located. By entering the command <EXP1> the user activated the application and was presented with a graphic display requesting the user to enter his initials and session number. This information became the labels for the data files created while the session was in progress. An instruction appeared instructing the user to press the "NEXT" key to begin the session. The "NEXT" key was a relabeled number 1 key on the numeric key pad. When this key was pressed the application software and signal set for that session was loaded.

All subjects were given the same signal set, containing both clean and noisy signals, for any given session. However, the order of presentation of the signals across sessions was randomized. The clean signals were never altered in any way. A different random number seed was used to create the randomized noise for the set of noisy signals in each session. This made it difficult for the user and the networks to learn the complete set. The variation in the noise reduced the dependence the users placed on the overall performance of the networks.

Once the signal set was loaded, the first pair of screens in the experiment were displayed. The text screen displayed what is illustrated in Figure 3.4A. The graphics screen automatically displayed the time domain of the first signal, as illustrated in the upper left corner of Figures 3-2 and 3-3. (Signals were presented to the test subjects in random order, but each session presented the same order across subjects). It was important to display the time domain plot automatically to avoid having users depend solely on the results of the neural networks as the only guidance for making their classification decisions. To minimize any potential influence on the decision-making process of the test subject, the time domain signal was the only mandatory tool presented in any experiment.

The subject could then use the space bar to hear the signal or press any one of five function keys to invoke the use of other tools or system controls. For instance, the user could press F2 (labeled ZOOM) to window in on a portion of the time domain signal. If this control were invoked, the user was required to enter two values representing the starting and ending points of the signal to be displayed (in the range of 0-499). The user then had to press the F1 key (TIME) to redisplay the time domain signal, this time seeing only the selected portion of the signal. Any time the "zoom" function was used, the frequency domain plot was redrawn to match the points displayed in the time domain plot. Pressing the space bar at this point audibly played the portion of the signal selected in the previous operation. The users could redefine the portion of the signal as many times as they wished. Pressing the F3 or "Frequency" key invoked the display of the frequency plot of the signal in the upper right corner of the screen, as seen in Figures 3-2 and 3-3. Pressing the F4 key produced a display of the spectrogram of the the signal. If the subject wanted to use the spectrogram in isolation of the other tools, he could do so, with the exception of the display of the time domain signal which was automatic. Using the windowing tool described above had no effect on the spectrogram plot. The complete graphics screen including all domain plots is illustrated in Figures 3-2 and 3-3.

5.3 Experiment 2: Networks Operating Alone

The purpose of the second experiment was to train and test networks to determine their capacity to perform acoustic classification. Previous experience with networks applied to acoustics problems provided direction and a methodology for determining the best types of networks to explore for this project. As part of the work ARD conducted for the Naval Air Systems Command (NAVAIR) on a similar neural network project, a software system was developed to allow efficient training and testing of a large number of networks. The system was used on this contract to develop the network configuration for the third experiment. The system and how the networks were trained are described below.

5.3.1 Network Training System

The network training system was designed to allow the operator to specify the parameters for several runs, each of which might take from several minutes to several hours. The specified networks could then be run consecutively without further input from the operator. This allowed for overnight runs of the networks, which did not interfere with the normal research activities during the day. Since the networks were run consecutively, it was necessary to devise the means to stop them after they had learned the task and before overtraining occurred.

Two methods were used to end a training run. No run was allowed to exceed a maximum number of iterations, but if the network had reached its state of best performance the run was stopped before the maximum number of iterations. Classification performance is cumbersome to test as often as necessary during a training run. Therefore, the measure of performance used during training was not how well the network classified the entire signal set, but its Mean Squared Error (MSE). The MSE is a measure of the difference of the desired output of the network from the actual output of the network. It is more restrictive than simply whether or not the class is correct.

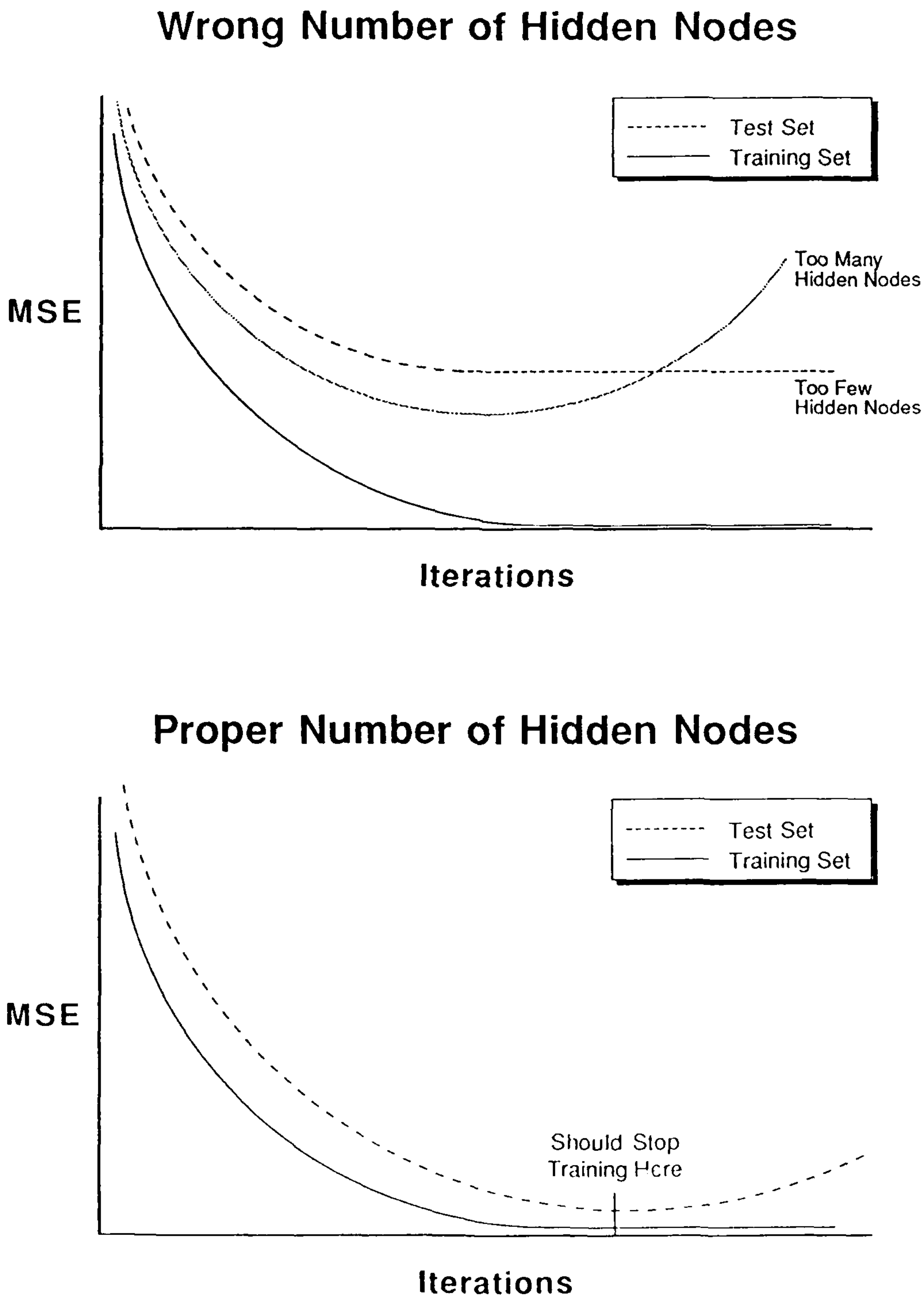
Since the network is attempting to produce known values at the output nodes for a given set of input signals, it is possible to measure the difference between the desired output for each signal and the actual output. This MSE is measured for two sets of signals: the normal training set, and a specially formulated set called the testing set. At regular intervals during the training run, training is disabled while the training set and the testing set are passed through the network and the MSE is calculated for both. At these intervals a copy of the network's weight structure is saved, in case the MSE of the network increases from this point forward.

It is typical for the MSE of the training set to asymptotically approach a minimum for any given number of hidden nodes. More training will continue to reduce the training set MSE towards the minimum. However, the real power of the neural network lies in its ability to classify signals outside the training set. This test set consists of signals from the same classes as the training set signals, but not identical to them. Performance on the test set, not the training set, is the proper measure of network performance. MSE performance on the test set does not generally follow the pattern of the training set. Instead, MSE of the test set usually reaches a minimum and then increases as training goes on, while the training set MSE continues to decrease. Too much training (beyond the global minimum for the MSE of the test set) will usually lead to poor performance on signals outside the training set.

The MSE on the test set is measured in the same manner and at the same intervals as on the training set. When the MSE on the test set stops falling and begins increasing, training is halted (see Figure 5-1) and the state of the network at the minimum testing MSE is recovered. The training runs typically went to a few hundred thousand iterations, although in many cases the significant training took place in the first few tens of thousands of iterations.

5.3.2 Training the Networks

Originally, ARD intended to construct a hierarchical set of three networks to process one parameter at a time. The first network would identify only one parameter. The second network would assume the results of the first network to be correct, and only concern itself with a subset of the signals. The third network would perform similarly, but have a smaller subset to deal with. After preliminary tests, it became clear that if the system made a mistake in the first level of processing, there was little or no hope that further processing could recover. In fact, it was likely that further processing on the part of the networks would only confuse the operator and degrade the overall performance of the system. For this reason, the hierarchical approach was abandoned in favor of using a single network to classify all aspects of the signal in a single pass. In this way, a network's low confidence in a single parameter would not forfeit the classification of the other parameters. No loss of performance was feared because the preliminary networks had shown reliable classification of all three parameters simultaneously was as well as for only one parameter.



MSE = Mean Square Error

Figure 5-1 BPN Topology Experimentation

The goal then became to find the best network capable of distinguishing the three parameters of interest: shell thickness, interior content and angle of insonification. The preliminary tests' dictation that all three parameters be included in the network led to fixing the number of hidden layer nodes at eight. Results from a previous project helped to fix several other network parameters. All possible networks had in common their input size and type, learning and smoothing rates, and output layer size. The input layer consisted of 500 nodes for time domain signals. The learning and smoothing rates used were 0.3 and 0.5, respectively. Lastly, eight output nodes were required, one for each specific parameter value. The input and output specifications are described in detail below.

The networks were trained on clean and noisy versions of the first and second averaged signals created from 16 of the 32 original instances of each signal. The signals were the same as those used in Experiment One, with the exception that no ramp was applied to the network signals. The third averaged signal for each class was used as the test set for measuring MSE, as described above.

Based on network performance and ease of adding noise, the time domain form of the signals was chosen as the prefered input type. The time domain form of the signals consists of 500 amplitude points in the range (-1, 1). This is the effective range of input values for the backpropagation network due to the transfer function of the nodes. The goal in making this transformation was to use the greatest range possible in the transformed values, thereby maximizing the differences between the signals and making the network's task easier. The same format was used in the human experiments, with the addition of ramping. Refer to Section 2 for a complete description of the signals.

A straightforward structure was selected for the output node results. One output node is assigned to each parameter of interest. For example, if a network were trained only to differentiate signals into 5 percent (thin) or 10 percent (thick) shell thicknesses, the network would have two output nodes. One node would be assigned to thin signals and one to thick. During training the thin output node would be taught to produce a high value (0.99) if the incoming signal is thin, while the thick output node produces a low value (0.01). If the signal is thick, the thick output node is taught to produce a high value while the thin output node gives a low value. When training is complete and the network is not told the class of the incoming signal, the activation on the output nodes determines the class of the signal. Whichever output node is higher is considered to be the estimate of the network. The networks trained here have eight output nodes: thick and thin; air, solid, and water filled; and 90, 45, and 0 degrees azimuth. The basic network architecture for all networks trained is illustrated in Figure 5-2.

5.3.3 Training Signals

The training involved running several networks to determine which performed best. Although many of the network parameters were unchangeable, as described above, two major factors were varied. The random initialization of the weights between nodes was changed because initial weight values have an impact on the final solution reached during training. Most importantly, however, the input signals were used in both clean and noisy form. The clean signals were simply the normalized, standardized 500 point signals described in Section 2. The networks trained with noise are described in Section 5.3.4.

The networks trained on the clean signal performed perfectly when tested against the clean version of the third averaged signal. To test the classification performance of the cleanly trained networks more thoroughly, the networks were tested against the third averaged signal at several levels of noise. The creation of the noisy signals and methods of testing against noise are described below.

5.3.4 The Effects of Noise

Since it had proven relatively easy to train a network to be a perfect classifier of the clean signals, the more difficult case of classifying under noisy conditions was evaluated. The signals used for training and testing contained a very small level of noise, as evidenced by the result of averaging eight signals in each class. This level of noise was clearly not difficult for the networks to handle. To test the networks under more difficult conditions,

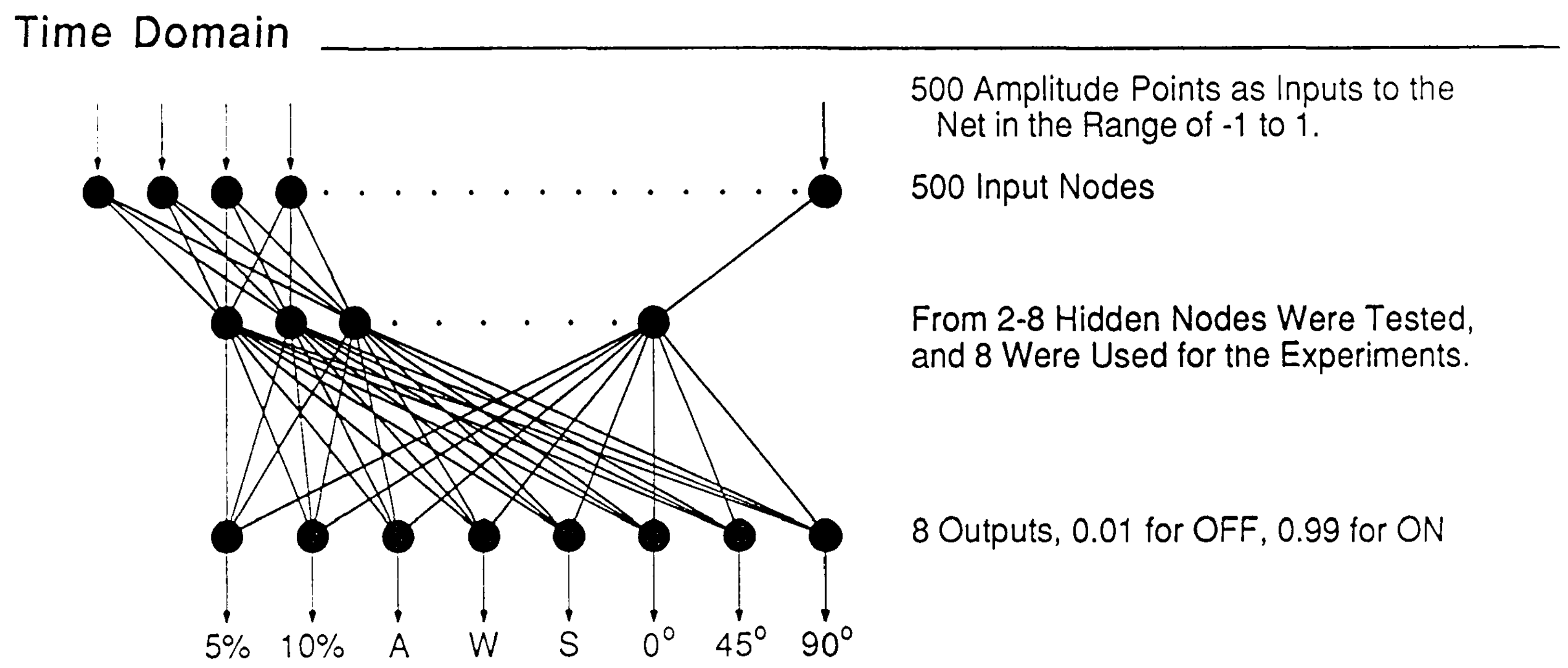


Figure 5-2 Network Architecture Used for Experiments 2 and 3

and to help determine the level of noise to use in the human testing, the networks were tested against noisy input. The noisy signals were generated by adding random noise sequences to the averaged signals of each class. In particular, random sequences were generated from a normal distribution with a mean of zero and standard deviation of 0.3. For each class of signal, eight levels of noise were used. Eight signal-to-noise ratios were computed using the formula

$$SNR = 20 * \log (\text{highest value in signal} / \text{standard deviation of noise})$$

$$= 20 * \log (\text{signal scaling factor} / 0.3)$$

The resulting signal-to-noise ratios were:

Signal Scaling Factor	SNR
1.6	14.6
1.4	13.4
1.2	12.0
1.0	10.5
0.8	8.5
0.6	6.0
0.4	2.5
0.2	-3.5

Each of the 18 averaged signals was multiplied by the eight different signal-to-noise scaling factors, and each of the eight resulting signals was added to one of the normal distribution random sequences. This produced complete sets of training signals with eight different signal-to-noise ratios. Appendix B shows A590 (Air Filled, 5% shell thickness at 90 degrees) in its averaged form and at two of the noise levels resulting from this process (8.5 dB and -3.5 dB).

During testing 20 signals at each of the eight noise levels for each class were used. Each of the 20 instances of the signals used a different random sequence, generated with the Microsoft C random number generator. However, the same random number sequence was used for a given instance across classes of signals. This prevented differences in the noise from affecting the results across all classes of signals.

The networks tested against these noisy signals had been trained on the clean versions of the first two averaged signals of each class. Those training signals had only a small amount of noise present. A hypothesis about network training states that when faced with substantial noise on the training signals, a network will be forced to derive any systematic information only from elements of the signal which will not be affected by the noise. If so, a network trained on noisy signals may be better equipped to handle noisy test signals. To test this hypothesis, networks were trained to classify content, thickness and angle using noisy signals. Two networks were trained at each of the eight noise levels using a new random sequence each time a training signal was needed by the network.

Figure 5-3 shows the classification performance of two networks, one was trained on averaged signals and the other was the best performing network trained on noisy signals. The test signals are at all eight noise levels, plus clean signals on which the network was not trained (these are labeled "infinite" SNR). In both cases classification performance shows a gradual decline as the noise level increases. There are no precipitous drops in performance as noise increases, and performance is still above chance (1/18) at the lowest SNR tested.

The network trained with signals at 8.5 dB SNR stayed above 90% correct classification until the SNR of the test signals was reduced to 6.0 dB or less. At higher noise levels performance degrades gradually. The training noise level must be raised to surprisingly high levels before the network cannot be trained to a good performance level. Adding noise to the training signals increased classification performance on noisy signals by very large amounts, and a fairly high level of noise on the training signals seems to produce the best results. This result has significant and positive implications for the ability of this technology to transfer to real-world situations with high noise levels.

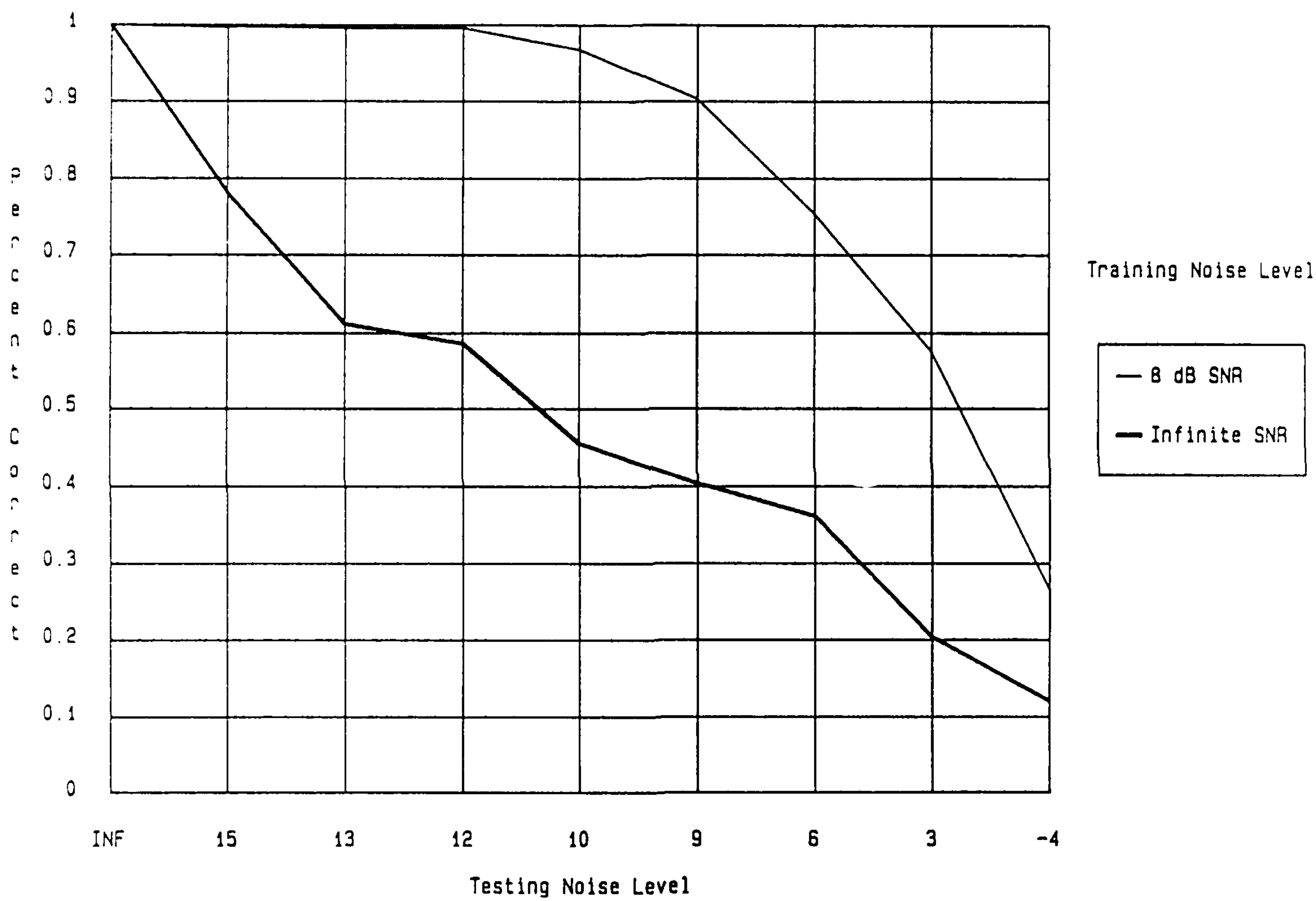


Figure 5-3 Classification Performance of two Neural Networks, One Trained on Clean Signals and One Trained on Noisy Signals, Both Tested Against Signals at Various Noise Levels

ARD

5.4 Experiment 3: Operators Using Base Tools and Neural Networks

Experiment 3 was carried out in the same manner as Experiment 1 with the important addition of the neural network classifier as a tool to be used in making personal judgements as to the three classification parameters. By pressing the F5 key, the user invoked the network display in the lower left corner of the screen shown in Figure 3.2. This display showed how the network had classified the three parameters and its confidence in each parameter's rating. The confidence bar graph was only labeled from high to low in order to reduce the user's dependence on the network results. It was also scaled in such a way as to rarely reach the high end of the scale.

6.0 RESULTS

6.1 Introduction

Upon completion of the experiments, ten subjects had run ten sessions each of Experiment 1 and five sessions each of Experiment 3. Experiment 1 evaluated the performance of humans without the aid of the network, and Experiment 3 examined human performance when the network was available. The performance of the neural network by itself was shown in two ways. First, Experiment 2 evaluated the network's performance against signals at a range of SNRs. These results are described in Section 5.3.4, and shown in Figure 5-3. Second, for the purpose of comparison with Experiments 1 and 3 the network's responses to the signals used in Experiment 3 were recorded and are used in the following analyses. All ten subjects' performances over all fifteen experimental sessions are shown in Appendix C.

The performance data were collected during the course of the experiment as the subjects made their choices for each signal. Their classifications were recorded by parameter (thickness, content, and angle) and for the signal as a whole. Classification of the entire signal is referred to as "overall" classification. As these data were collected the frequency by which the subjects used each of the tools was also recorded.

Performance varied greatly among the ten subjects. In particular two of the subjects showed much stronger performance against the noisy signals than the rest of the group. These two subjects are singled out at one point in the analysis to compare network performance with the best human performance. A repeated measures analysis of variance (ANOVA) procedure was applied in several ways to these performance data to discover the statistically significant effects of the experiment. The performance data were summed over each session to give the number of correct classifications, by parameter and overall, for the session. Subsets of this data set were created to analyze different aspects of the experiments. These analyses are presented below.

The use of the tools by the subjects in different situations is also of great interest. A correlation analysis is done for each of the human experiments to determine statistically significant relationships between the use of tools and the performance of the subjects.

6.2 Training Effects

The first analysis is concerned with training effects, increases in performance as the sessions progressed, under clean and noisy conditions. The performance results of each subject, by session (1-10 of Experiment 1) and by noise level (Noisy or Clean) were submitted to ANOVA. The performance of the subjects over the ten sessions is shown in Figure 6-1. Both clean and noisy signals are classified with increasing accuracy over the course of the experiment, with noisy signals more difficult to classify. The effect of noise on performance is significant, $F(1,9)=6.84$, $p<.05$. The training effect of the sessions is also significant, $F(9,81)=5.73$, $p<.001$. This demonstrates that subjects improved in the task over the time allotted for the experiments, and that noisy signals provide a significantly greater challenge than clean signals. The average number correct advanced from 2.3 to 9.1 for clean signals and from 1.1 to 5.4 for noisy signals. There is no significant interaction between noise level and session. The large variability from subject to subject in performance on clean signals is somewhat surprising. It also appears that classification performance is still increasing at the end of the experiment. A longer test in which the subjects are allowed to reach asymptotic performance would better test human-network interaction. This should be carried out in future studies.

6.3 Training Effects by Parameter

The second analysis looked at the same training effects across noise conditions, this time by parameter instead of overall. That is, for each subject, session, and noise level the performance on each of the three parameters is reported separately. Since thickness is chosen from only two possibilities instead of three for content and angle, these values are scaled down to 2/3 of their original values to make the chance values of these parameters equivalent. This is only done in the case of human subjects, for

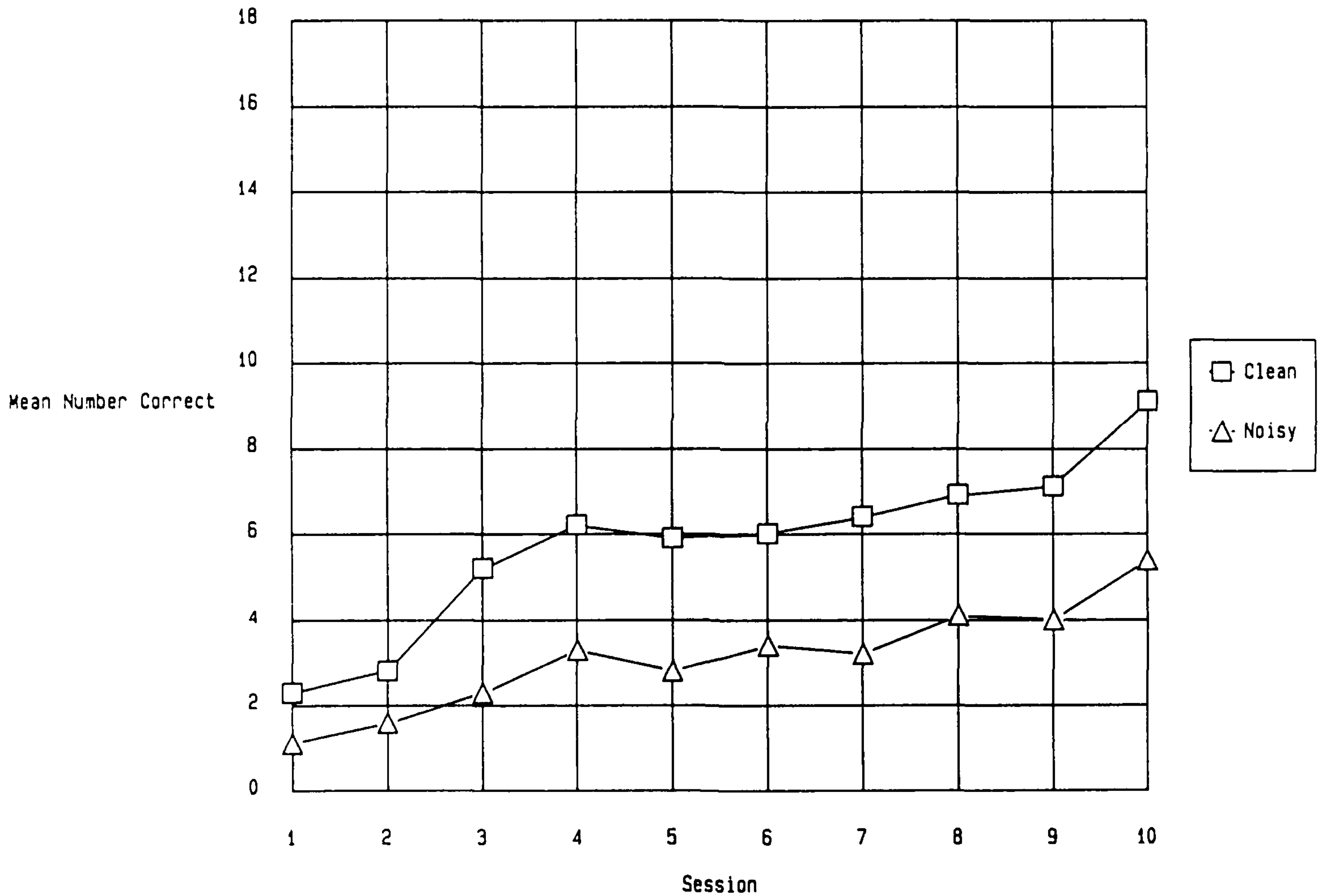


Figure 6-1 Classification Performance of Subjects in Experiment 1, for Clean and Noisy Signals

ARD

whom performance values are not close to the maximum possible values. Figure 6-2 shows performance averaged over the three parameters for each session, by noise level. As in the first analysis, there is a significant training effect, $F(9,81)=6.07$, $p<.001$, and a significant noise effect, $F(1,9)=9.20$, $p>.025$, but no significant interaction between noise and session. The average number correct advanced from 7.14 to 11.66 for clean signals and from 6.31 to 9.50 for noisy signals.

Figure 6-3 shows performance as a function of parameter, averaged over all subjects and sessions, for clean signals, noisy signals, and an average of both. The significant effect of parameter is clear here, $F(2,18)=46.93$, $p<.001$. The performance difference between angle judgements and thickness judgements is apparent, with thickness performance barely above chance levels. Angle proved the easiest parameter for the subjects to judge. For clean signals, an average of 7.11 correct thickness judgements were made per session, while 12.94 correct angle judgements were made. This is almost certainly due to the distinct shape of the 90 degree (broadside) waveforms, which have a strong initial specular return followed by little remaining energy. This is in sharp contrast to the 45 and 0 degree signals. Content performance falls somewhere between thickness and angle performance.

Further inspection of Figure 6-3 reveals a relatively large effect of noise on content and angle judgments, but relatively little effect on noise on thickness judgments. This is revealed in a statistically reliable noise by parameter interaction, $F(2,18) = 5.48$, $p < .025$, which most likely reflects a floor effect in the thickness judgement.

A significant interaction between parameter and session, $F(18,162)=2.82$, $p<.001$, is also due to the difficulty most subjects experienced in classifying thickness. As shown in Figure 6-4, there is very little improvement in the number of correct thickness judgements averaged over subjects. Angle judgement shows quick improvement early, and content judgement shows similar but smaller improvements. These differences produce the interaction effect. There is no significant three-way interaction between parameter, noise, and session.

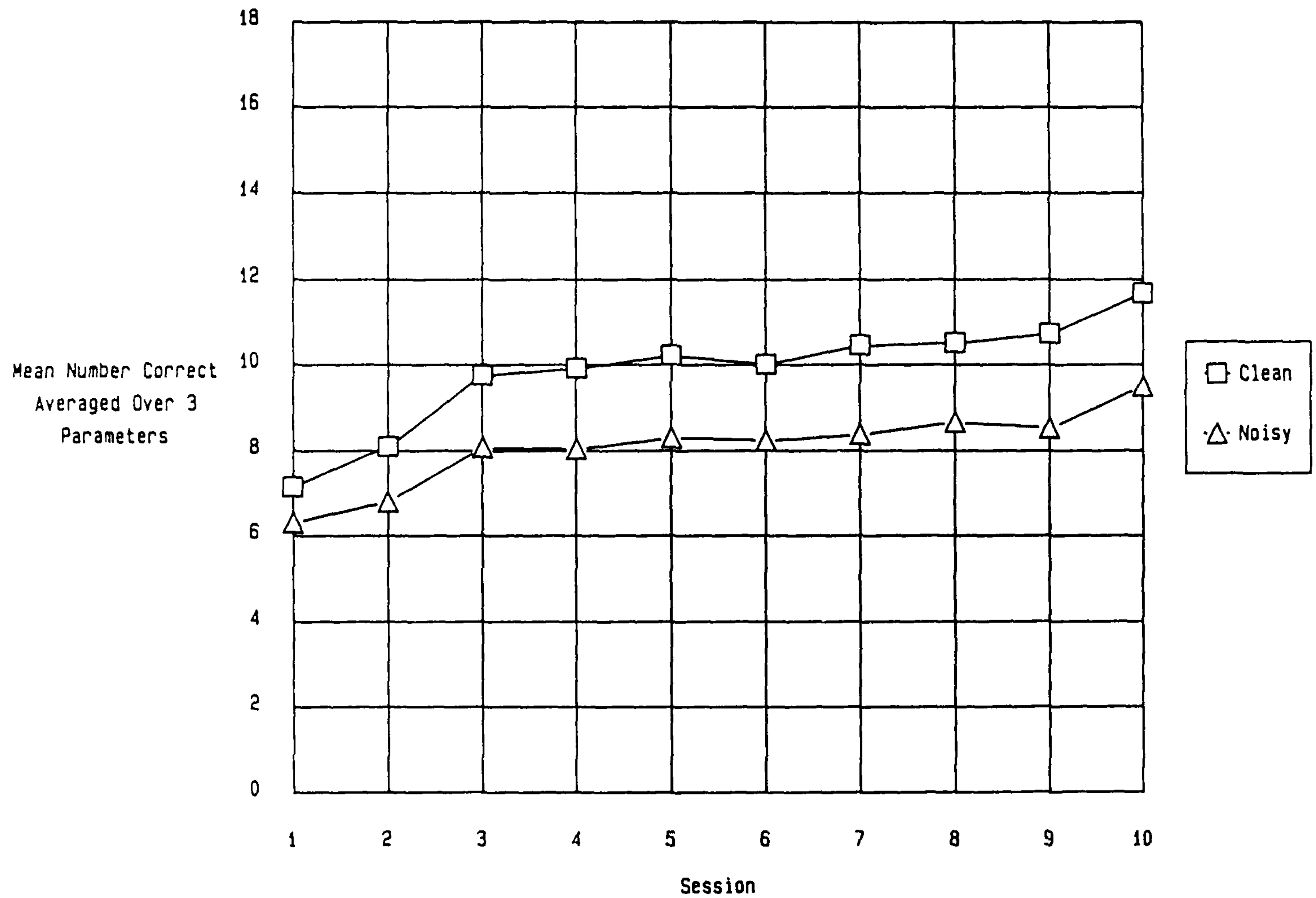


Figure 6-2 Noise Effect on Classification of the Three Parameters in Experiment 1

ARD

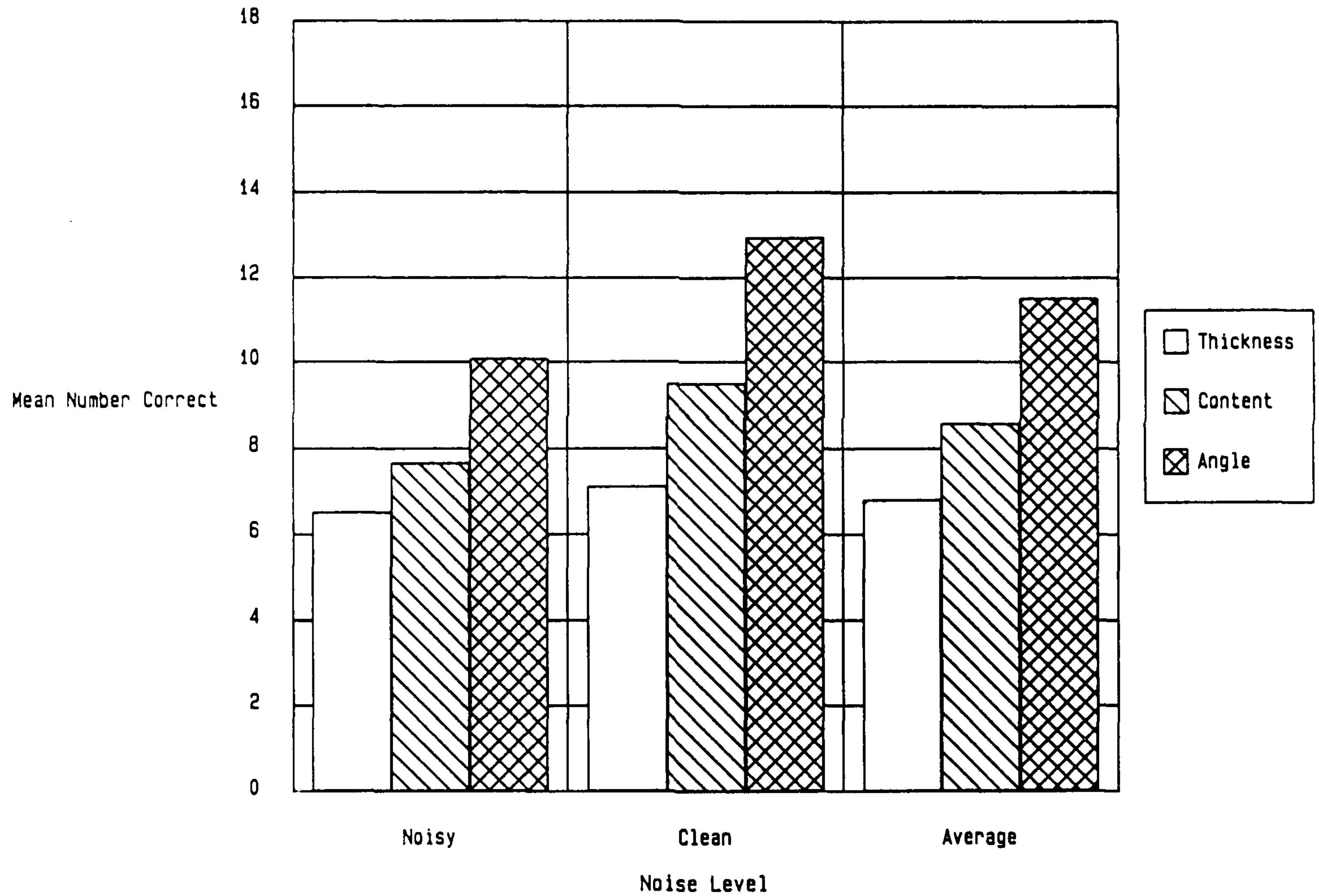


Figure 6-3 Classification Performance in Experiment 1 by Parameter, Averaged Over Ten Sessions

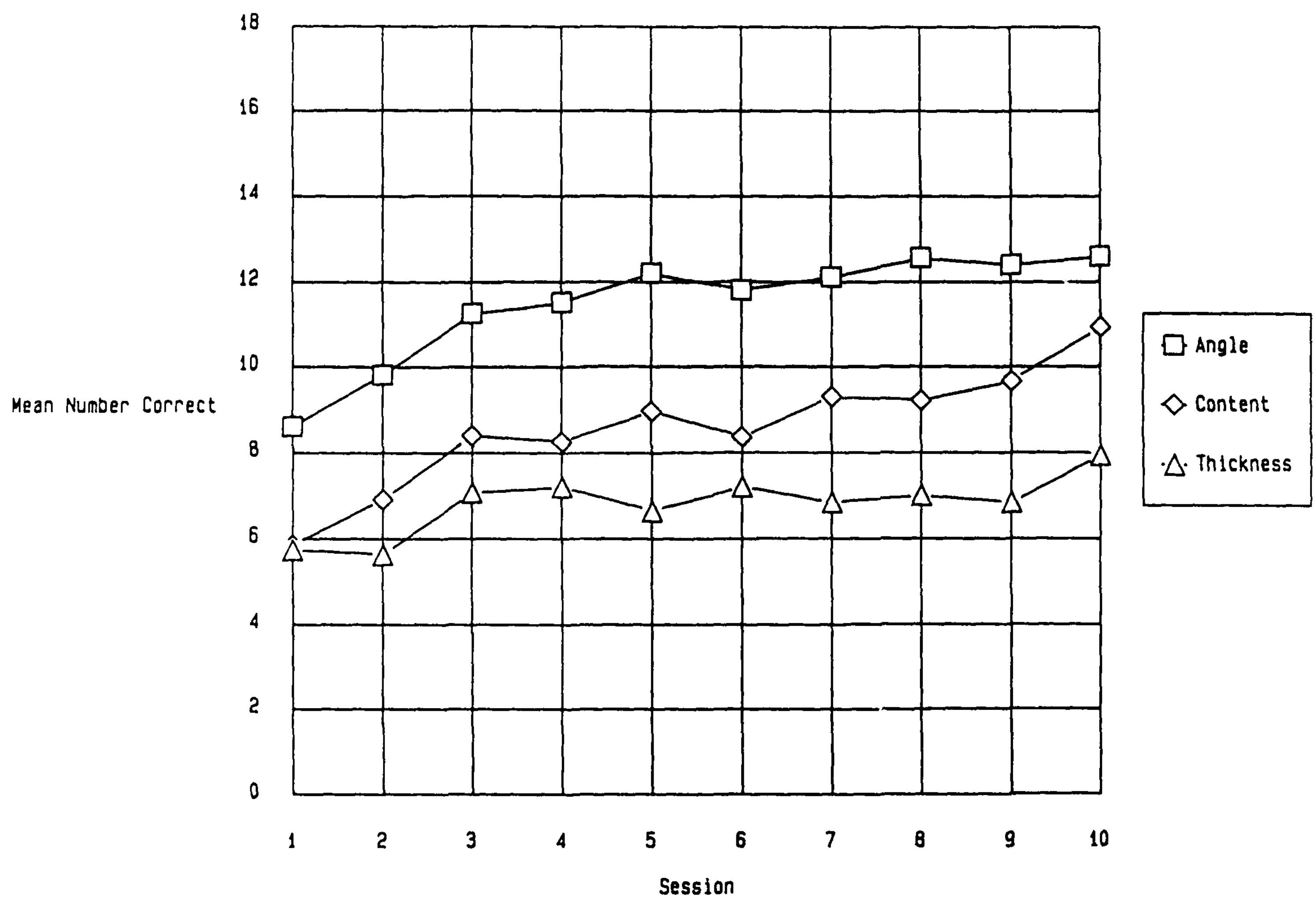


Figure 6-4 Classification Performance in Experiment 1 by Session for Each Parameter, Averaged Between Clean and Noisy Signals

6.4 Effect of Neural Network as a Tool

In the third ANOVA analysis, the effect of having a neural network classifier available is considered. The overall performance data were arranged by subject, session, the presence of a neural network (Yes or No), and noise level. Since the subjects used the network for five sessions in Experiment 3 only the last five sessions of the subject-only data are considered in this analysis. By these sessions the subjects are assumed to have learned most of what they will learn over the ten sessions. Figure 6-5 presents this data.

Session effects have been studied in earlier analyses, and the significant noise effect, $F(1,9)=18.90$, $p<.005$, is expected from previous results. The very large effect of the network is of primary importance, $F(1,9)=65.47$, $p<.001$. Previous results showed the excellent classification performance of the network alone, most importantly in noisy signals, and it is not surprising that the subjects as a group performed much better with the network available than without. On clean signals, the subjects averaged 7.1 correct classifications without the network and 17.0 with. On noisy signals the average number of correct classifications rose from 4.02 to 15.22. The subjects quickly learned that the network was better at classifying the signals than they were.

6.5 Performance of Network Alone

To further characterize the performance of the networks, the fourth analysis arranged the network's overall performance by noise level. These were the data for the network acting by itself to classify the same signal set that the subjects classified in Experiment 3. These data are shown in Figure 6-6. Classification of clean signals is perfect, 18 correct in each session, while the average number of noisy signals correctly classified is 16.4 over the five sessions. Two sessions were perfect, one recorded 16 correctly classified noisy signals, and two had 15 correctly classified noisy signals. The difference in classification performance between noisy and clean signals is not significant, $F(1,4)=5.57$, although the number of incorrect responses came very close to the intended level of ten percent. At the higher noise levels needed to reduce network performance further, human performance is expected to drop

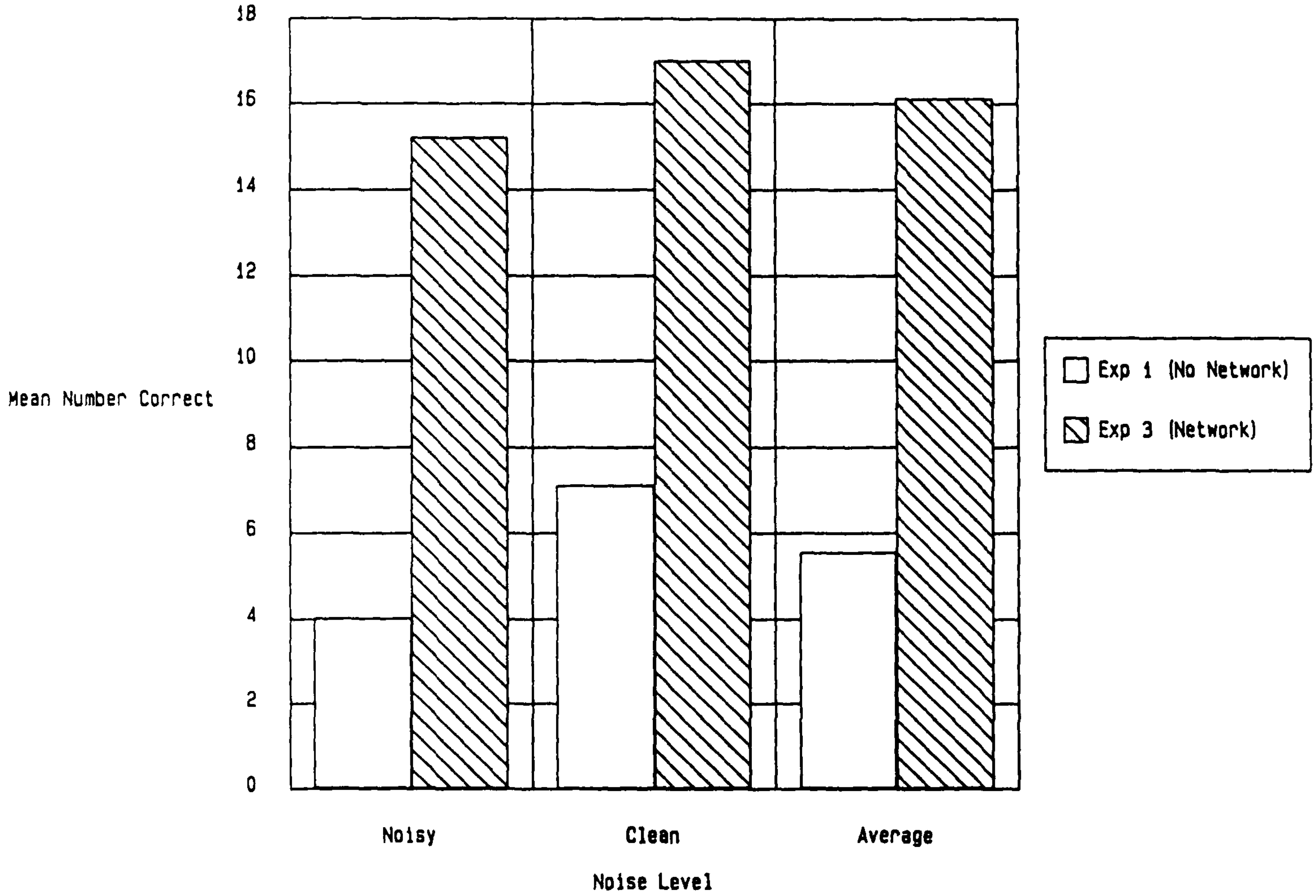


Figure 6-5 Overall Classification Performance of Subjects in Experiment 1 vs. Experiment 3

ARD

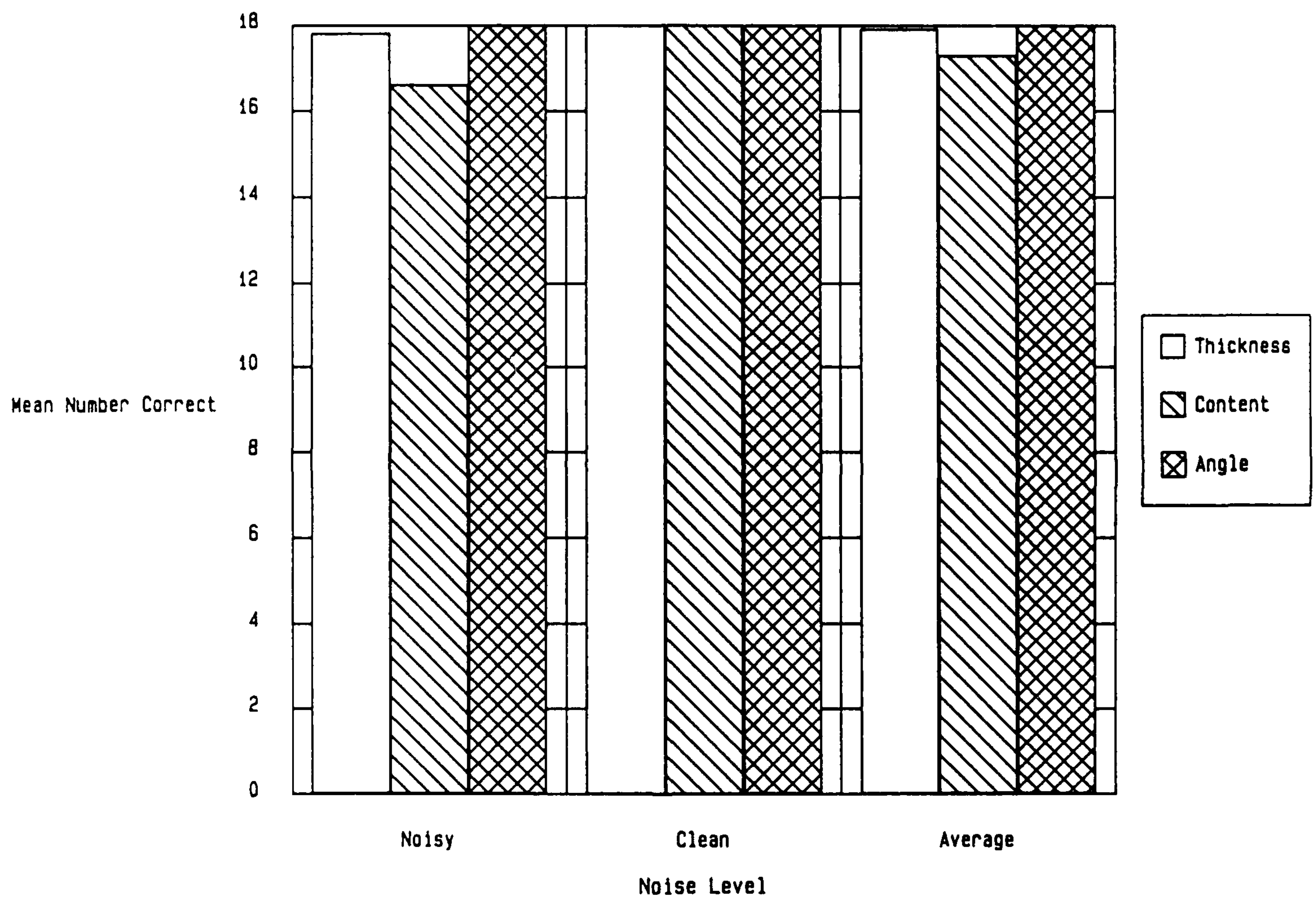


Figure 6-6 Classification Performance of the Neural Network, by Parameter, on the Five Signal Sets Used in Experiment Three

precipitously. The effect of parameter is significant, $F(2,8) = 4.65$, $p<.05$. This is due to the drop in performance on content, the only parameter the network had significant trouble with when the signals were noisy. The average number of noisy signals correctly classified on content was 16.6. The thickness of noisy signals was correctly classified an average of 17.8 times per session, and angle was classified perfectly.

6.6 Comparison of Subjects and the Network

Having established the individual performance of humans and networks on the given signal set, and of humans acting with networks, it remains to compare the performances of all three conditions. For this purpose, the final ANOVA concerned the overall performance of humans without networks (using the last five sessions of Experiment 1), humans with networks (using the five sessions of Experiment 3), and networks alone (using the network's response to the signals of Experiment 3). For the two cases in which subjects were involved, the average performance over the subjects was used since there exists only one network "subject" against which they were compared. These data were arranged by session, by "classifier" (Human, Human with Network, and Network), and by noise level.

Two of the subjects markedly outperformed the group. Poorly performing subjects might be expected to follow the judgement of the network, which the subjects could see performing well during Experiment 3, without much additional effort to improve on the network's performance. The two excellent subjects are expected to have the best chance of improving on the network's performance. For this reason two analyses were done, the first using an average of only the two top performers and the second using an average of all ten subjects.

6.7 Network Use by the Best Two Subjects

Figure 6-7 shows the results of the top two performers versus the network. A significant noise effect is apparent, $F(1,4)=63.36$, $p<.005$, as is a significant effect of "classifier" (human, human and network, network), $F(2,8)=38.64$, $p<.001$. There is also a significant interaction between the two, $F(2,8)=21.71$, $p<.001$. The effect of noise is easy to attribute primarily to the subjects

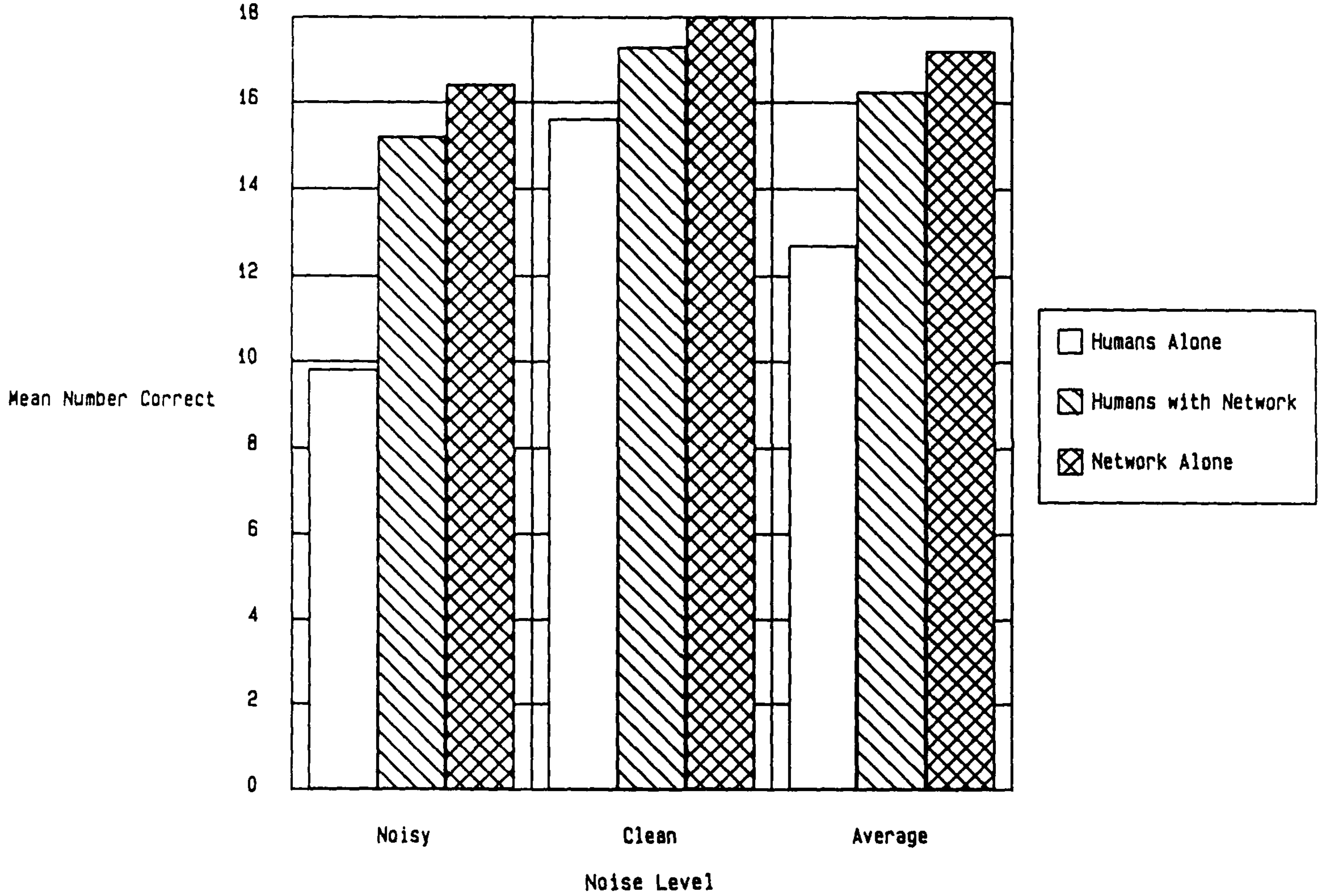


Figure 6-7 Classification Performance of the Two Best Subjects, With and Without Network Aid, and Networks Acting Alone

ARD

acting alone in Experiment 1. There the two subjects showed much greater performance on clean than noisy signals (15.6 correct averaged over session, versus 9.8, respectively). Much smaller differences are apparent for the other two conditions. These two subjects showed a smaller performance difference between noisy and clean signals than most of the subjects.

The significant effect of "classifier" is clearly due to the relatively poor performance of the subjects acting alone compared to the two conditions in which networks were involved. The subjects acting alone averaged 12.7 correct responses per session, while the network alone averaged 17.2 correct responses per session. While this is not surprising, the performance of subjects acting with the aid of networks did not exceed that of networks acting alone. The ultimate goal of such a system is to take advantage of the best aspects of humans and networks to form a system superior to either alone. A posthoc analysis by a series of t-tests reveals that most differences between the performances of the "classifiers" in noisy and clean conditions are significant. Only for clean signals is there no significant difference between the two subjects acting alone and with the aid of the network. Nor is there a significant difference between the two subjects using the network and the network acting alone. There is, however, a significant difference between the subjects alone and the networks alone.

Since the network is known to be perfect on clean signals, the only explanation for less than perfect performance of the subjects in Experiment 3 for clean signals is that they sometimes disagreed with the network, incorrectly. When noise is introduced, however, the two top subjects are unable to perform better with the network available as a tool than the network could on its own.

The interaction of noise and "classifier" is again due to the relatively poor performance of the subjects acting alone on noisy signals. While the three "classifiers" showed similar performance on clean signals, the two top subjects acting alone dropped in performance much faster than the two conditions in which networks were involved, when noisy signals are considered.

6.8 Network use by the Entire Group of Subjects

Figure 6-8 shows the second analysis, in which the human data is an average of all ten subjects rather than the high performing subset of the subjects. The only apparently substantial change is in the performance of humans acting alone in Experiment 1. Their scores drop to 7.10 on clean signals and 4.02 on noisy signals. There is a significant effect of noise, $F(1,4)=28.75$, $p<.01$, as before. Since the difference between the low performance of humans alone and the high performance when networks are involved seems to be the reason for significant effects of the experimental condition, it is not surprising to find a significant effect of "classifier" (humans, humans with network, network) here, $F(2,8)=307.06$, $p<.001$, and a significant interaction between noise and "classifier", $F(2,8)=7.76$, $p<.025$. These are the same effects seen when only the two top subjects are used, with the lower human performance making these effects more pronounced.

The same series of posthoc t-tests as was performed on the two subjects was applied to the ten subjects. In this case the difference between the subjects' performance on clean signals with and without the network is significant, as is the difference between the subjects' performance on clean signals with the network and the network's performance alone.

6.9 Tool Use

To analyze the patterns of tool use by the subjects the frequency of use of the tools was correlated with classification performance. The last five sessions of Experiment 1 and each of the five sessions of Experiment 3 were used. For each subject, the use of each of the tools was totaled over each session. To this list was attached the number of correct classifications by parameter, and overall. Separate data were compiled for Experiments 1 and 3. Correlations were taken for each experiment separately. In addition, the same analysis was done for the two top performing subjects alone to see if they used the tools in any different manner than the subjects as a whole. Significant correlations are assumed at the .05 level.

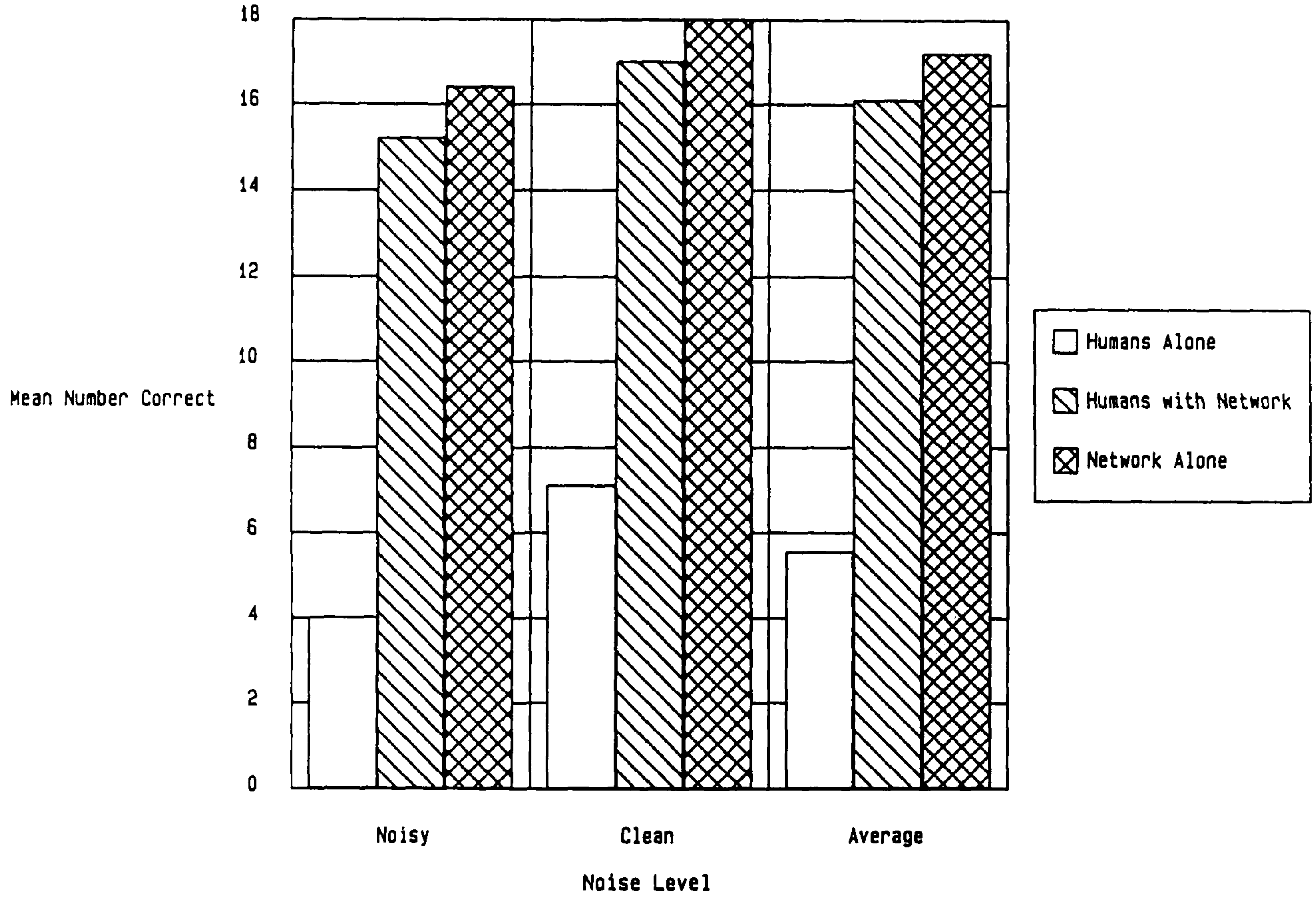


Figure 6-8 Classification Performance of all Ten Subjects, With and Without Network Aid, and Networks Acting Alone

6.9.1 Tool Use: Humans Alone

The only significant correlation (at the .05 level) for the ten subjects classifying the clean signals is between playing the signal and angle classification performance, and this is a negative correlation. It is hypothesized that the ten subjects applied fairly different strategies to the initial task, resulting in few correlations across session.

When faced with noisy signals, subjects seem to rely on the various graphical tools more. The use of frequency display, and of the combination of frequency and spectrogram, are correlated with overall classification performance. When these tools were used, performance was higher. Subjects tended to do better on thickness and content when these displays were used more frequently. In contrast, use of the spectrogram tool is marginally significantly correlated (.10) to performance on angle. The frequency of playing the noisy signals is negatively correlated with overall classification performance.

It is expected that playing the signal is the most familiar tool available to these novice subjects. They may rely on it earlier in the test and come to understand the other tools over time, while their scores are improving from remaining practice effects. This would account for the frequency of playing the signals decreasing while performance increases.

6.9.2 Tool Use: With Network Aid

Again there are few correlations other than the significant negative relationship between overall performance and frequency of playing the signal. Since one may expect practice effects to be leveling off by these last five sessions, this relationship might be due to uncertainty over the more difficult signals. Faced with the harder signals the subjects may be trying to gather more information by playing the signal more frequently. Given that the network is prone to failure on only some of the noisy signals, the subjects may be playing these signals more to make up for shortfalls of the network.

6.9.3 Tool Use by Top Two Subjects

Some marked changes occurred when only the top two performing subjects were included in the correlation. These subjects showed a higher number of significant correlations under all conditions than the entire group of subjects did. This suggests that the more these two subjects used the group of tools, the better they were able to classify. In Experiment 1 the two subjects showed the same significant negative correlation between the frequency of playing the signal and overall performance that the ten subjects showed. However, in Experiment 3 this reversed to a significant positive correlation. This appears to be due to the strategy of one subject to refrain from playing the signal late in Experiment 3, presumably when the subject felt knowledgable about the other tools based on previous excellent performance. The subject then fell slightly in performance without the information provided by playing the clean signals.

7.0 DISCUSSION

Several major conclusions can be drawn from the results reported in the previous section. First, as expected, both session and signal noise had a major effect on the classification performance of human listeners. Second, differences occurred in classification performance across the three individual signal parameters and these differences appeared to change with subject experience. Third, the use of classification tools developed in relatively complex ways with experience and the pattern of use differed substantially across the individual subjects. Fourth, the artificial neural networks (ANNs) performed nearly perfectly as planned. These high levels of network performance led the human users to rely nearly exclusively on the networks as a decision aid to the neglect of the other tools. Here again, large differences occurred in this pattern across the individual participants. Each of these major conclusions is considered in more detail below and recommendations are drawn for further human research using the test-bed system developed on this project.

7.1 The Role of Practice and Signal Noise on Overall Performance

The classification task used here proved to be a difficult one for human subjects. This is clear from both the overall correct performance in which all three of the characteristics of the insonified objects are identified correctly as well as from performance on the individual parameters considered alone. On both measures, performance was shown to increase with practice and to degrade when simulated environmental noise (0 mean Gaussian noise) was added to the signals. Both results were anticipated, but a closer examination of individual differences revealed some interesting findings. For these and other finer-grained analyses we focus on the last five blocks of the first experiment after subjects were familiar with the task. Substantial individual differences occurred with overall correct performance with clean sounds ranging from 94% to 9% (mean of 39%) and from 56% to 9% with the noisy sounds (mean of 22%). The overall group was split into "good" and "bad" performers at the median level for clean signals. These two groups were differed in several ways, two of which are spelled out here. First, the good classifiers showed steady

ARD

improvement over the five sessions on both the clean (slope = 7.3%) and noisy signals (slope = 3.6%), whereas the bad classifiers improved little (slope = 0.3 and 1.6, respectively). Second, although performance for both groups suffered with the addition of noise (62% versus 32% and 17% versus 13%, for the good and bad groups, respectively), the impact of noise was far greater on the good performers than on the poor performers. Although even the bad subjects were performing above chance, this group difference likely reflects "floor effect" in the weak performers' data. It is also possible that the good classifiers simply chose to focus on the clean rather than noisy sounds, with the bad classifiers focusing on both. The exact reason for these individual differences cannot be determined on the basis of the present data. However, it seems clear that future experiments should incorporate a longer training period and perhaps a more careful screening and selection of subjects.

7.2 The Role of Object Parameters

The three physical parameters of the objects insonified to derive the test signals used here manifest themselves in different ways acoustically. For example, in our previous research we have shown object angle to be primarily a time-domain feature, whereas object thickness is primarily a frequency-domain feature and contents incorporate both time- and frequency-domain characteristics. As mentioned in the previous analysis, large differences in performance occurred across these parameters. As in the overall analysis, the good classifiers did much better than the bad on the individual parameters (average correct = 80% and 52% for the two groups, respectively). Most interesting, however, is the relative difficulty experienced by the two groups. In particular, the good group identified all three parameters at well above chance levels (81%, 74% and 85% for thickness, contents and angle, respectively) whereas the bad group performed at chance on thickness (50% chance) and only slightly above chance on contents (33% chance) (45%, 42% and 70% for thickness, contents and angle, respectively). (Note that the data were adjusted to compensate for the differences in chance level in the ANOVA analyses reported in the previous section). This result suggests that the poor classifiers may have had particular difficulty in extracting frequency-domain information which, by our previous findings, should be especially important for thickness and somewhat important for contents judgments.

7.3 The Role of Signal Processing Tools

A major purpose of this study was to investigate the ability of novice users to use a range of classification tools which included, time-, frequency- and spectrographic-plots, an acoustic display and a zooming capability. (The ANNs as a decision tool will be considered separately below.) The study was successful in demonstrating the individuals can and do learn to use these tools over a relatively short training period. Several specific findings are noteworthy. First, the zooming capability was virtually never used by any of the subjects and will not be considered further. We still see this as a potentially useful tool to the analyst which should be retained in future studies involving more highly trained users.

Second, the acoustic display was widely used by almost all subjects. On the average, subjects listened to each sound a surprising 13.2 times on each trial over the final six sessions in the first experiment. Furthermore, individuals in the good group listened more frequently than those in the bad group for both clean (13.8 versus 9.8 per sound) and noisy (16.1 versus 14.7 per sound) sounds. Note also that subjects listened to the noisy sounds (14.7) more often than to the clean sounds (11.8). Interestingly, use of the acoustic display alone did not distinguish good from bad performers (for example, one of the single best subject rarely listened to the clean sounds). Rather, a more subtle pattern of tool use distinguishes good from poor listeners. This pattern will be considered below. Third, there were even greater group differences in the use of the visual aids. Specifically, the good performers used both the frequency and spectrographic displays on substantially more of the trials than did the poor performers (86% versus 44% and 86% versus 71% for the two displays, respectively). Moreover, the better subjects tended to use these displays together (76% of the trials) whereas the weaker subjects did not (43% of the trials). This pattern of tool use is consistent with the performance data described previously. In particular, our understanding of the acoustic cues important for distinguishing among the object parameters indicates that time-, frequency-, and time by frequency or spectrographic information will all be required. The time-domain plot of each signal is provided automatically on each trial, but the other displays must be requested explicitly. The good subjects appear to have learned this, whereas the weak subjects did not. Most notable is the comparatively limited use that poor subjects made of the frequency-domain display. Interestingly, this coincides with the great difficulty shown by these individuals with the frequency-based, thickness parameter. Similarly, the only moderate use of the spectrographic display coincides with the problems they experienced with the time/frequency, content parameter. Unfortunately, since the time-domain display was always provided, we cannot comment on its role in signal analysis. In future experiments more could be learned by providing no default information. In other words, users should be required to request all displays so that comprehensive tool use data could be obtained. In addition, it would also be instructive in future research to examine classification of single- as well as multiple-parameter sound catalogs. In the present study sounds which differed in all three parameters were included, and as pointed out above, optimal performance would likely involve all of the decision aids provided by the test-bed system. This strategy was adopted in order to obtain as much information as possible in the limited time available. By examining classification of selected subsets of the full catalog, selective tool use could also be investigated. For example, if only signatures from *objects of a single shell thickness* were presented, time-domain displays may become less important.

Fourth, although widely applied even on the final session, use of the acoustic display declined with practice for the good subjects while the graphical tools (frequency and spectrographic displays) increased in use with practice for these individuals. No discernible trends in tool use occurred in the poor subjects' data. We interpret these trends to reflect the more sophisticated analysis carried out by the stronger subjects. The acoustic display is obviously a more "natural" presentation of sound than are the graphical displays. Nonetheless, listening is not necessarily the most useful technique for an analyst. Hence, the better subjects listened less and used spectral analysis more as they improved, whereas the weak subjects continued to rely predominantly on listening. This pattern accounts for the negative correlations reported in the results section between listening and other tool use as well as between listening and performance. It is important to note, however, that this result is correlational--it does not suggest that increased listening leads to poor performance in the task.

Fifth, both the good and poor subjects used at least one of the visual aids more often on the noisy signals (96% and 81% of the trials, respectively) than on the clean signals (76% and 61% of the trials, respectively).

7.4 Artificial Neural Network Performance and Use

As expected from our previous research, the ANNs performed perfectly when clean signals were used. As indicated previously, we added noise to the signals in an effort to degrade the ANN's performance to a more realistic level. We faced a delicate trade-off here since human performance declined dramatically with noise levels of only moderate difficulty for the networks. For this reason, we selected a signal-to-noise ratio for our noisy sounds which had only a minimal impact on the network performance (approximately 10% decline). Even this conservative choice led to a major deterioration in listener performance.

The consequence of this disparity between the network and listener performance was a nearly complete reliance on the ANN tool for decision making. This tool was used by virtually every listener on nearly every trial for both the good and poor performers (90% and 96% of the trials, respectively). As described in the results section, this tool had a major impact on classification performance. Overall performance improved dramatically to 95% and 85% for the clean and noisy signals, respectively. This improvement occurred for both the stronger and weaker subjects (91% and 89%, respectively). Interestingly, most subjects continued to experiment with at least some of the other decision aids. Specifically, although there was a general decline in the frequency of listening after the ANNs were made available in the third experiment, this display was still widely used by both good (4.2 plays per sound) and poor (9.1 plays per sound) subjects. Similarly, the visual aids were used less often with the networks, but were still used by many subjects, especially for the noisy signals. These findings on the use of an ANN decision aid lead to at least two recommendations for further research. First, it would be of interest to examine the frequency of network use as its reliability is degraded. Clearly, this cannot be accomplished by adding noise to the signals, i.e., at the network input, since human performance would decline to chance levels. Rather, performance could be degraded by adding noise to the network outputs, hence, achieving the objective of introducing errors into the network output without damaging the signal quality. Second, it would also be interesting to examine the role that ANNs may play as a training tool. The observation that users continued to examine the signals both acoustically and visually while basing their judgment on the networks suggests that learning may be continuing. If subjects were retested without the network tool in subsequent sessions, the significance of ANNs as a training aid could be determined.

7.5 Summary

The first experiment established that people were capable of learning to classify this signal set. Performance did increase over the sessions, reaching a reasonably high level on average. A large amount of variability exists among the subjects, with some still apparently at chance levels while two were excellent at classifying clean signals and good at noisy signals. The use of the various tools available varied widely from subject to subject, indicating that such an analysis system is difficult to adapt to many operators if it depends on only one method of deriving information from the signals. People show strong differences in how they best interpret information, and a classification system which provides various means of interpreting a signal will find greater acceptance and higher performance from a population of users.

Experiment 2 showed that the neural network proved to be a much better classifier than the average of the subjects, and somewhat better than the best of the subjects, on clean signals. Experiments 1 and 3 were limited in duration, and we may expect further learning to take place in longer tests. This is certainly the case in real-world systems, in which parity might be expected between operator and network, and in many cases superiority of the operator. When presented with noisy signals, the network strongly outperformed the subjects. The learning curve on noisy signals may be very long for the subjects, but the capability of the network on noisy signals is outstanding. While the best two subjects classified noisy signals correctly more than twice as frequently as the average of all ten subjects, the network was far better.

When given the network as an additional tool in Experiment 3, the subjects soon came to depend on it. It so outperformed most of the subjects that they abdicated most decisions to it. When the subjects disagreed with the network, the subjects were usually wrong, resulting in slightly lower scores than the network itself. In the time allotted, the subjects were unable to identify the faults and strengths of the network so that they could know when to trust it and when to override it. This situation is likely to change when the system is faced with real-world signals of higher complexity

In conclusion, this study accomplished the primary objectives set forward in the introduction. We have demonstrated clearly that naive human users can learn to perform a demanding acoustic analysis task and to use a variety of decision aids in the process. Furthermore, the results described in this report make it clear that tool use depends on the interaction of a number of different factors. Some have been tentatively identified in this report and others must await further research. We conclude that the test-bed system developed here will be an extremely effective tool for understanding the complex dynamics of acoustic analysis.

8.0 BIBLIOGRAPHY

Gorman, R. P., & Sawarti, T. (1985). The use of multidimensional perceptual models in the selection of sonar echo features. The Journal of the acoustical Society of America, 77, 1178-1184.

Hickling, R. (1962). Analysis of echoes from a solid elastic sphere in water. The Journal of the Acoustical Society of America, 34, 1124-1137.

Howard, J. H., Jr., (1983). Auditory perception in loose parts monitoring, Washington, D.C.: U.S. Nuclear Regulatory Commission, (NUREG/CR-3008).

Howard, J.H., Jr. & Ballas, J.A. (1983). Perception of simulated propeller cavitation. Human Factors, 25, 643-656.

Minsky, M. & Papert, S. (1969). Perceptrons. Cambridge, MA: MIT Press.

Morse, Philip M. (1983). Vibration & Sound. American Institute of Physics, New York.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J., "Learning Internal Representations by Error Propagation", in: Rumelhart, D. E. & McClelland, J. L. (1986)(Eds.). Parallel Distributed Processing: Exploration in the Microstructure of Cognition pp. 318-362 Cambridge: MIT Press.

Stanton, T.K. (1989). Simple approximate formula for backscattering of sound by spherical and elongated objects. JASA, 86, 1499-1510.

Talamo, J. D. C. (1982). The perception of machinery indicator sounds. Ergonomics, 25, 41-51.

Vanderveer, N. J. (1980). Ecological acoustics: Human perception of environmental sounds. Dissertation Abstracts International, 40, 4543B.

Warren, W. H., Jr., & Verbrugge, R. R. (1984). Auditory perception of breaking and bouncing events: A case study in ecological acoustics. Journal of Experimental Psychology: Human Perception and Performance, 10, 704-712.

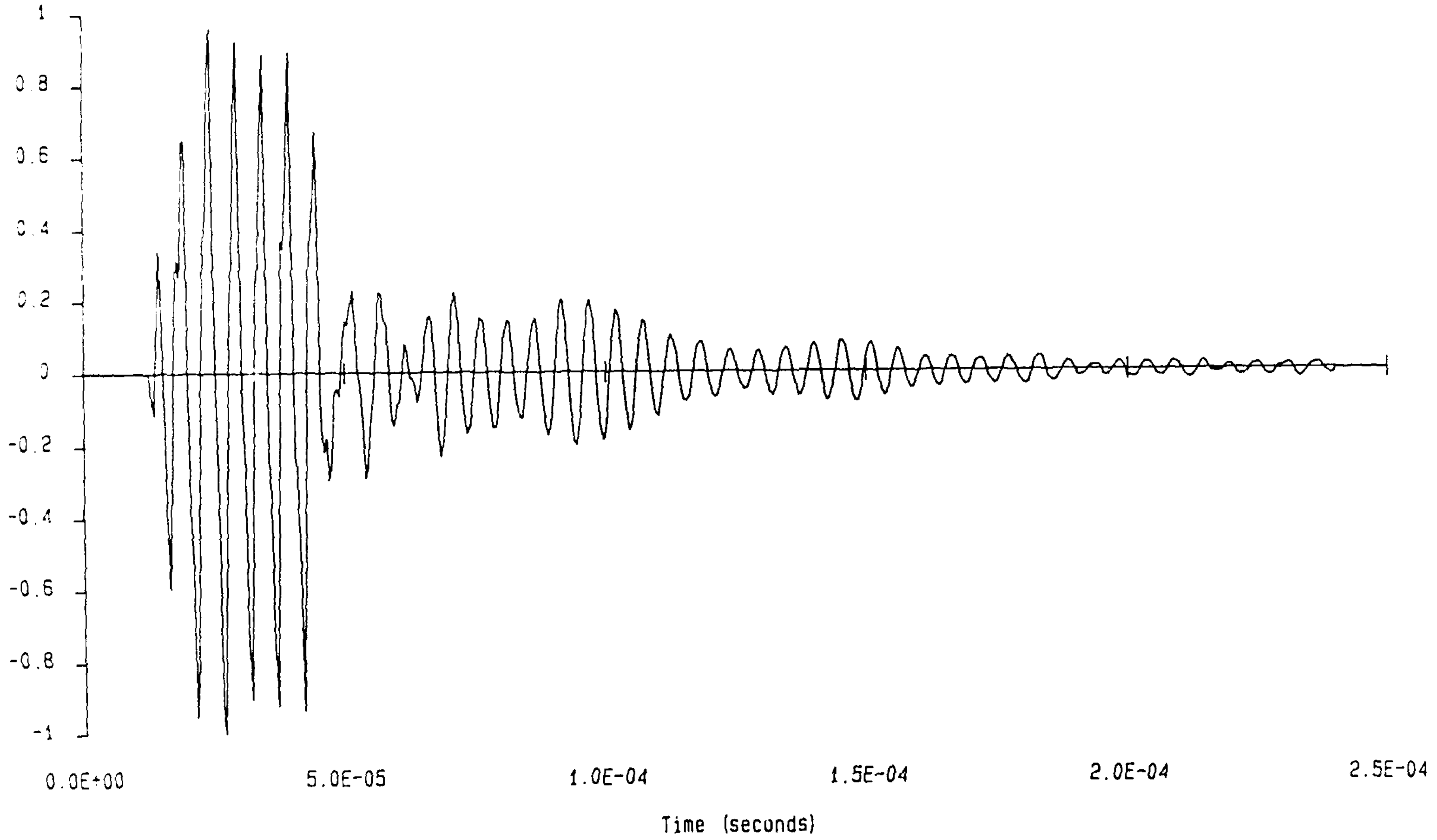
ARD

APPENDIX A

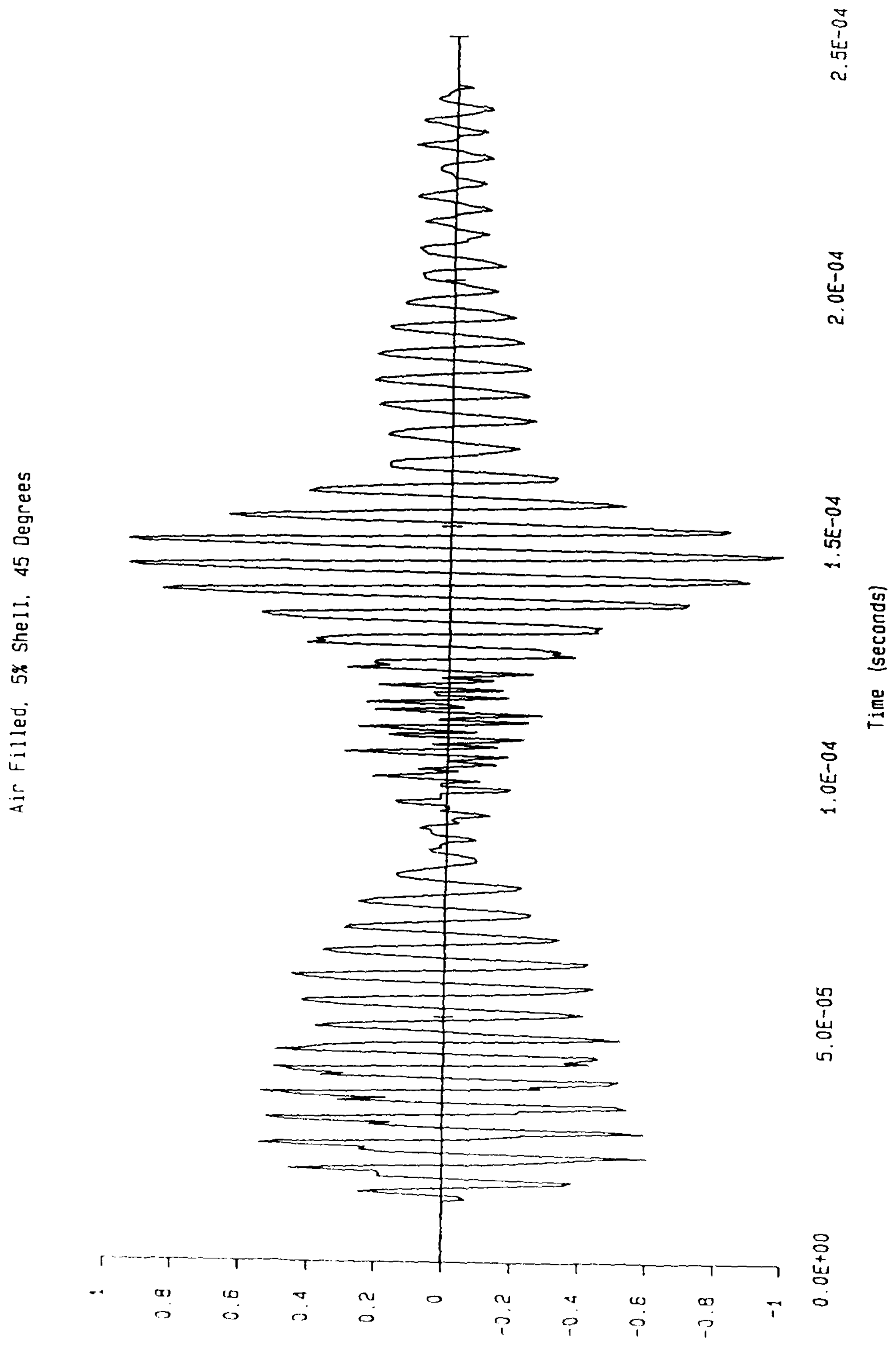
AVERAGED SIGNALS

These are the eighteen averaged signals used in much of the analysis as the clean version of the signals. Appendix B illustrates the effect of adding noise to these signals. Here the signals have been averaged across eight samples of the signal per class. The mean is shifted to zero, amplitude is normalized to the range (1,-1) and standardized to 500 points. See Sections 2.3 and 2.4 for further details.

Air Filled, 5% Shell, 90 Degrees

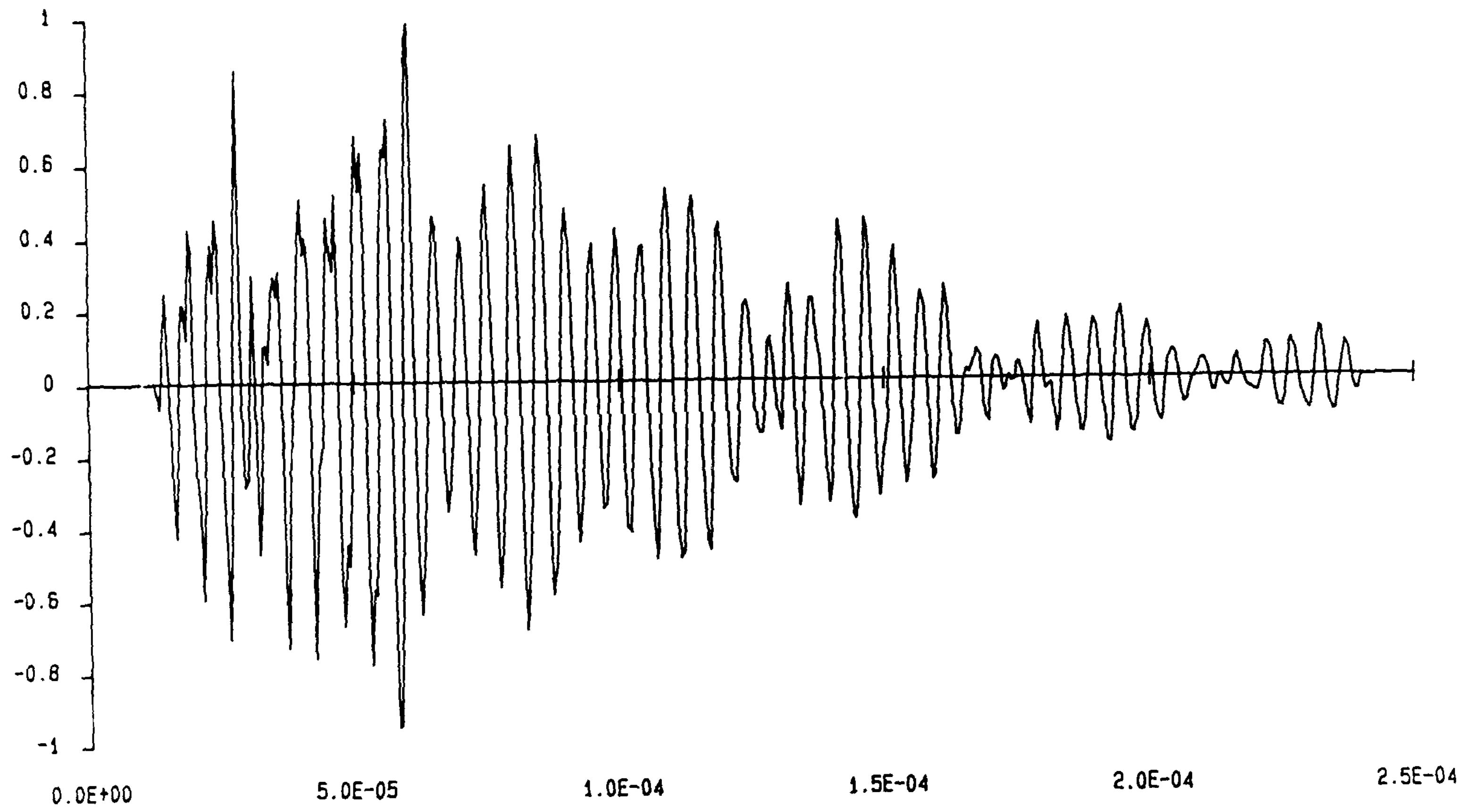


ARD



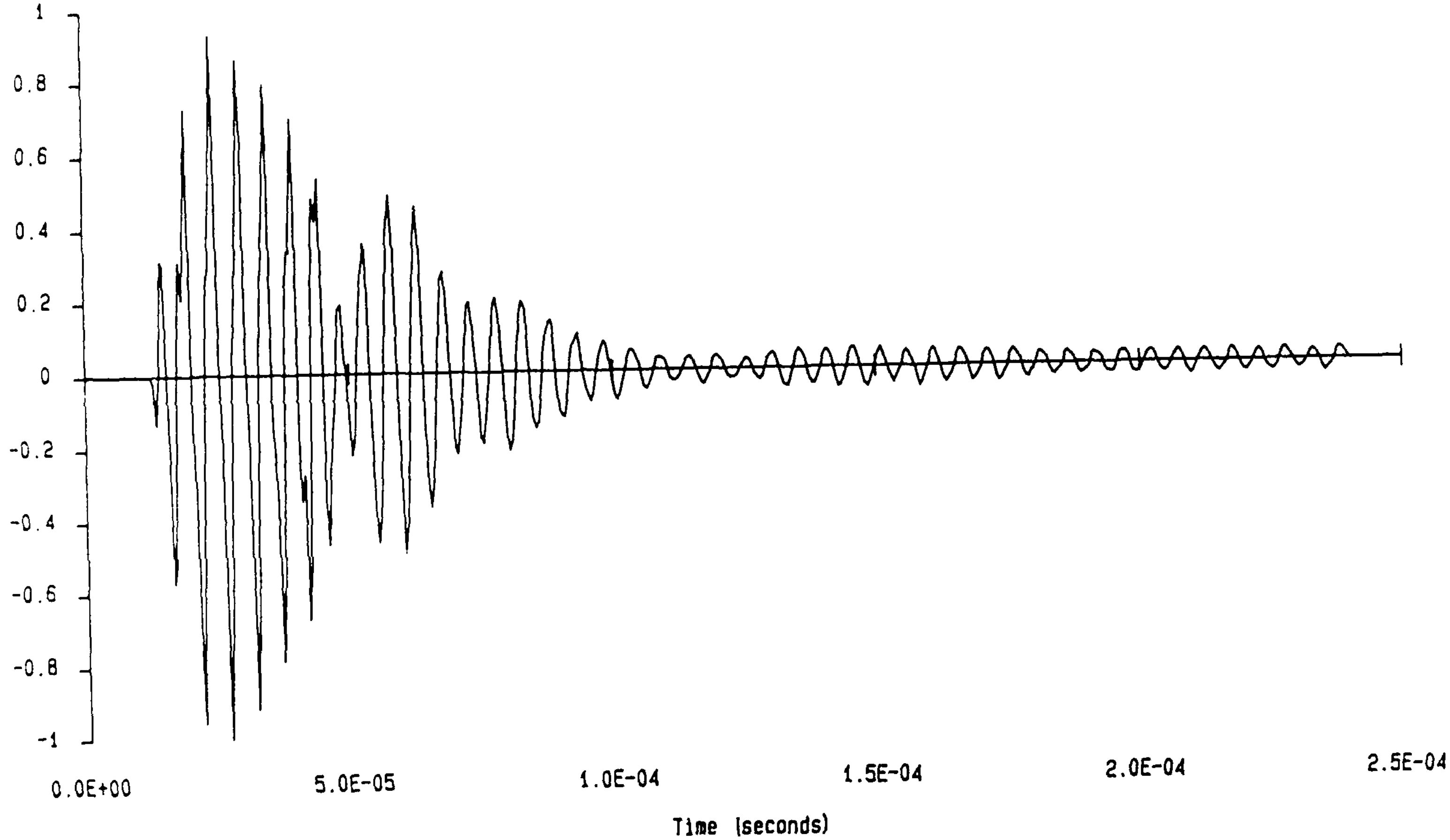
Air Filled, 5% Shell, 45 Degrees
Time (seconds)
0.0E+00
5.0E-05
1.0E-04
1.5E-04
2.0E-04
2.5E-04
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Air Filled, 5% Shell, 0 Degrees

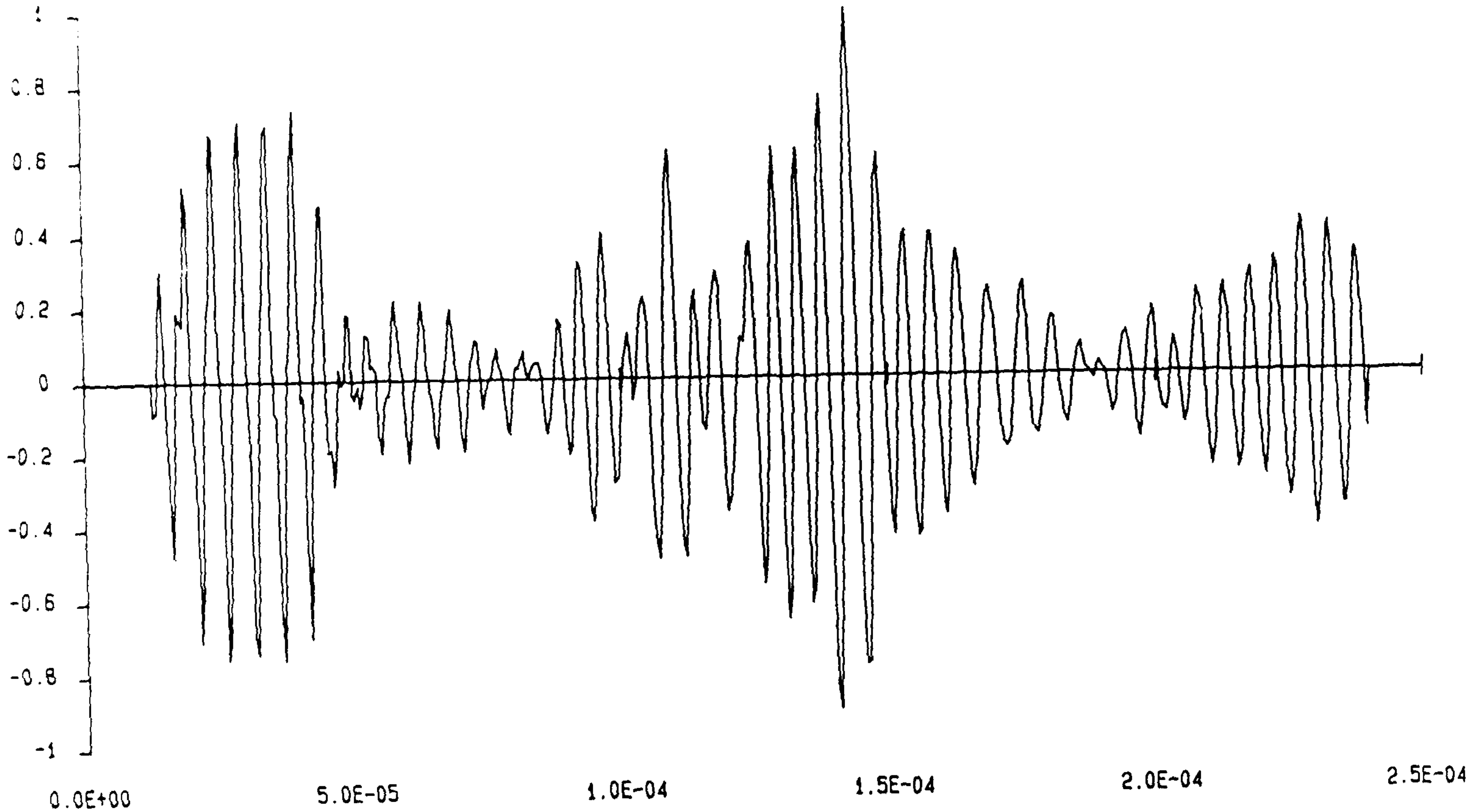


ARD

Air Filled, 10% Shell, 90 Degrees

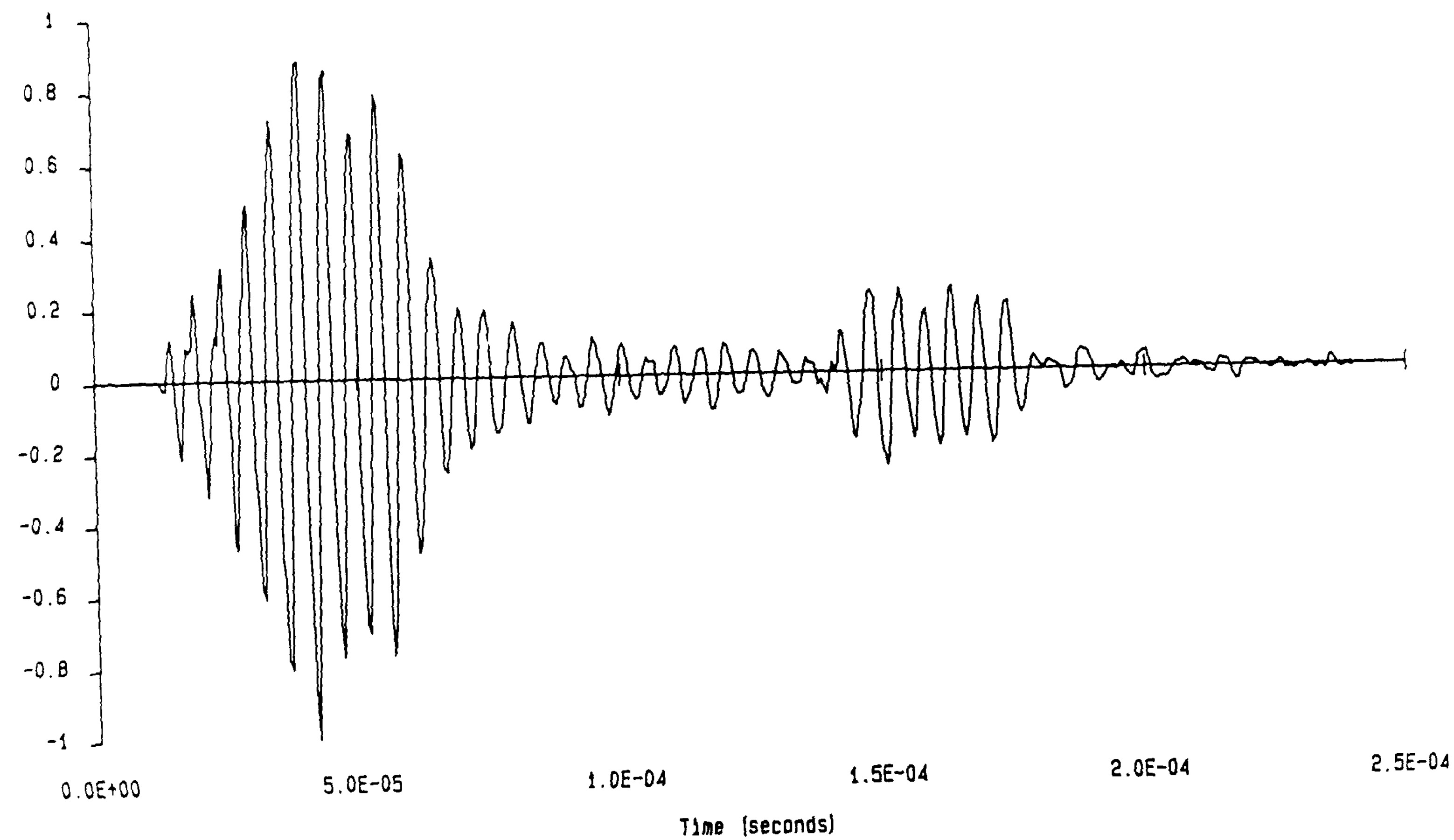


Air Filled, 10% Shell, 45 Degrees

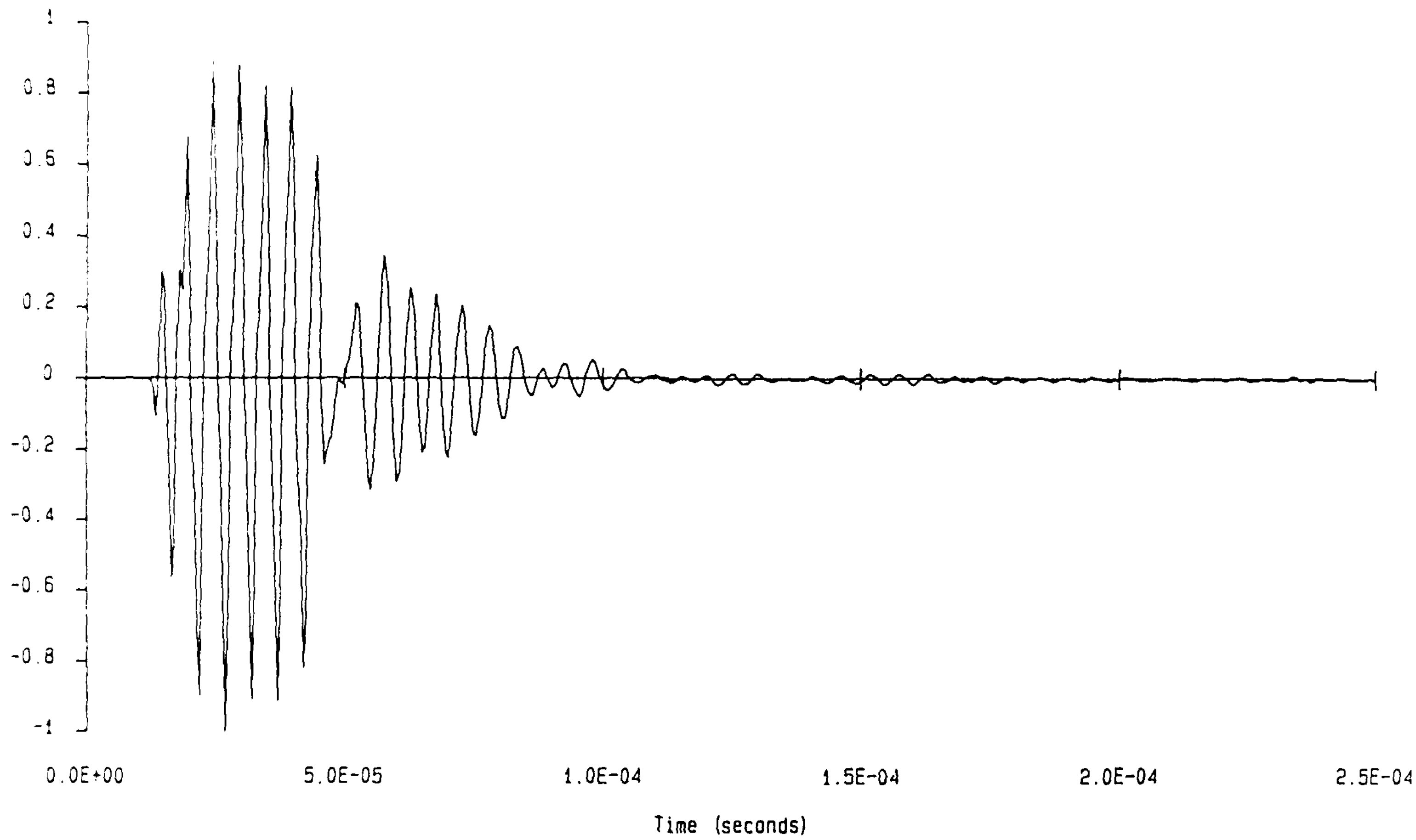


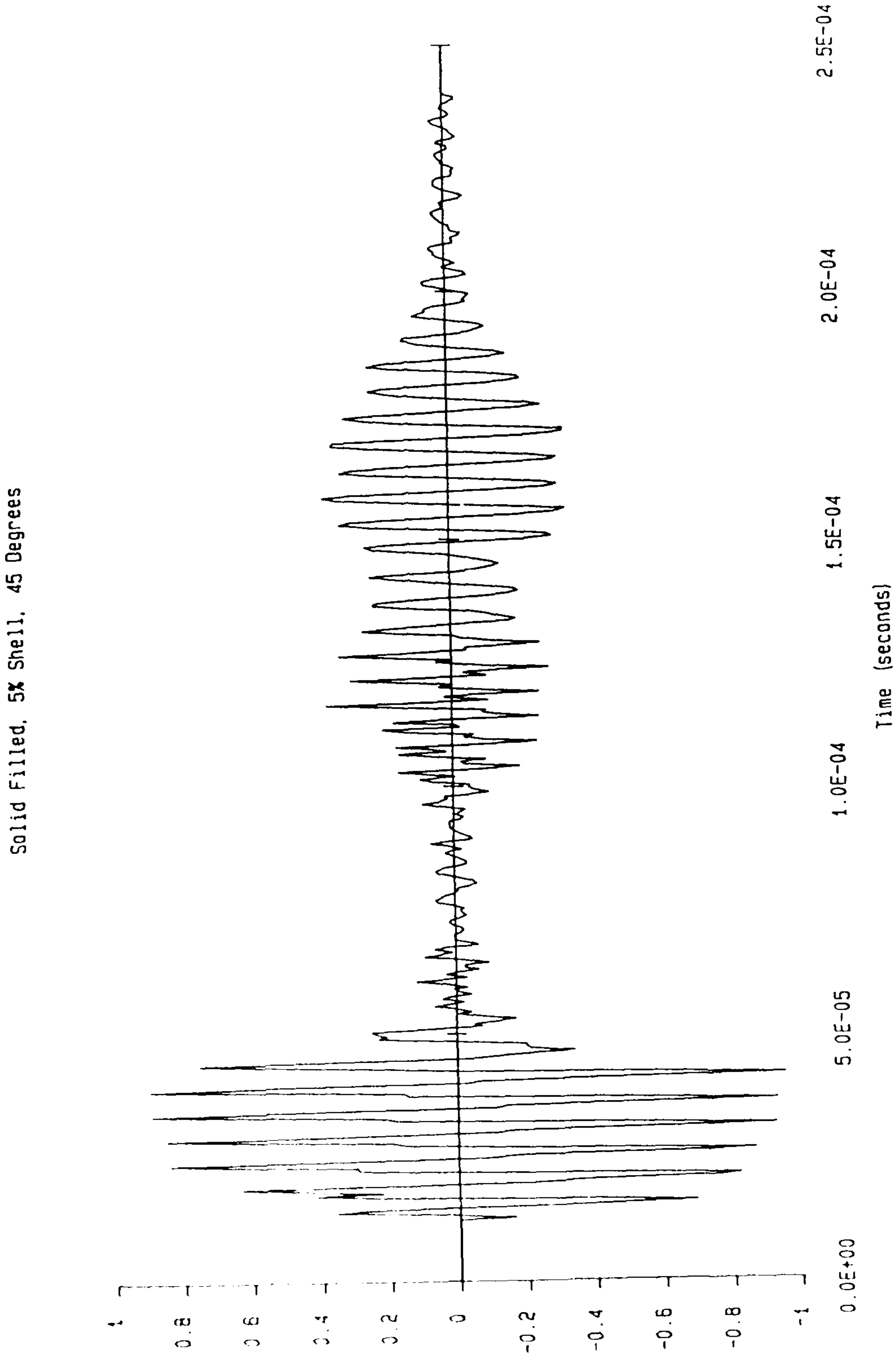
Time (seconds)

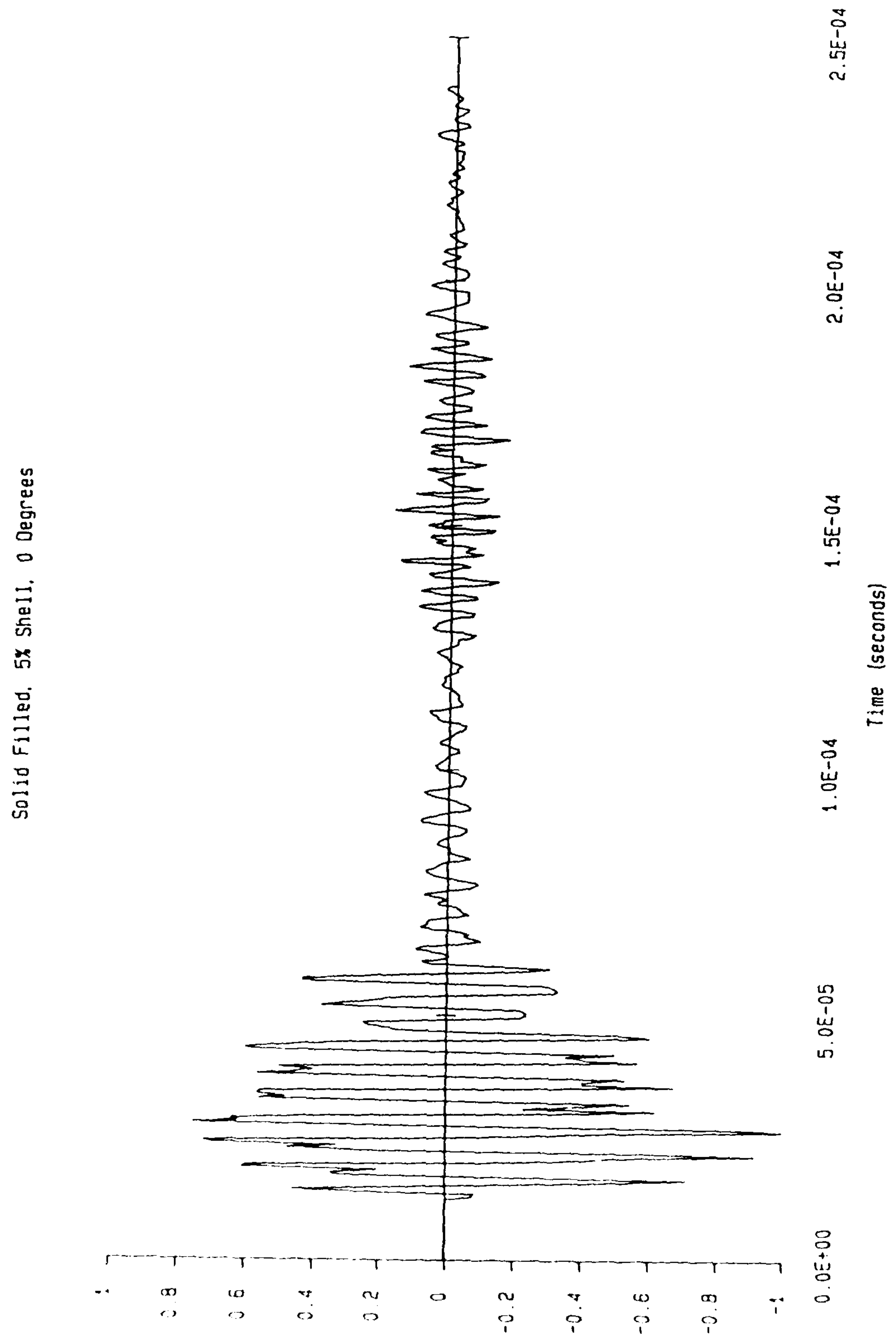
Air Filled, 10% Shell, 0 Degrees



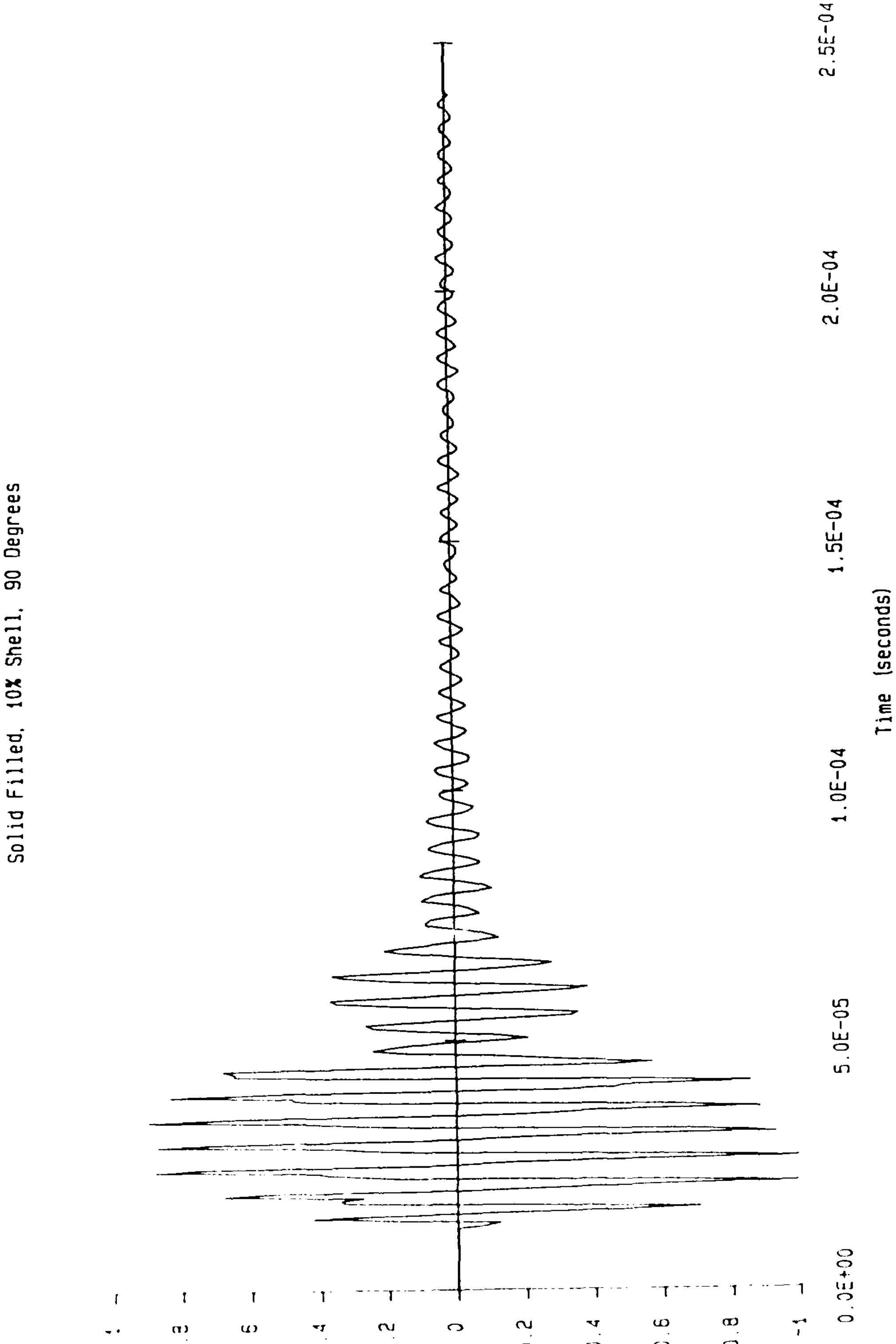
Solid Filled, 5% Shell, 90 Degrees




Solid Filled, 5% Shell, 45 Degrees
Time (seconds)
0.0E+00
5.0E-05
1.0E-04
1.5E-04
2.0E-04
2.5E-04
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

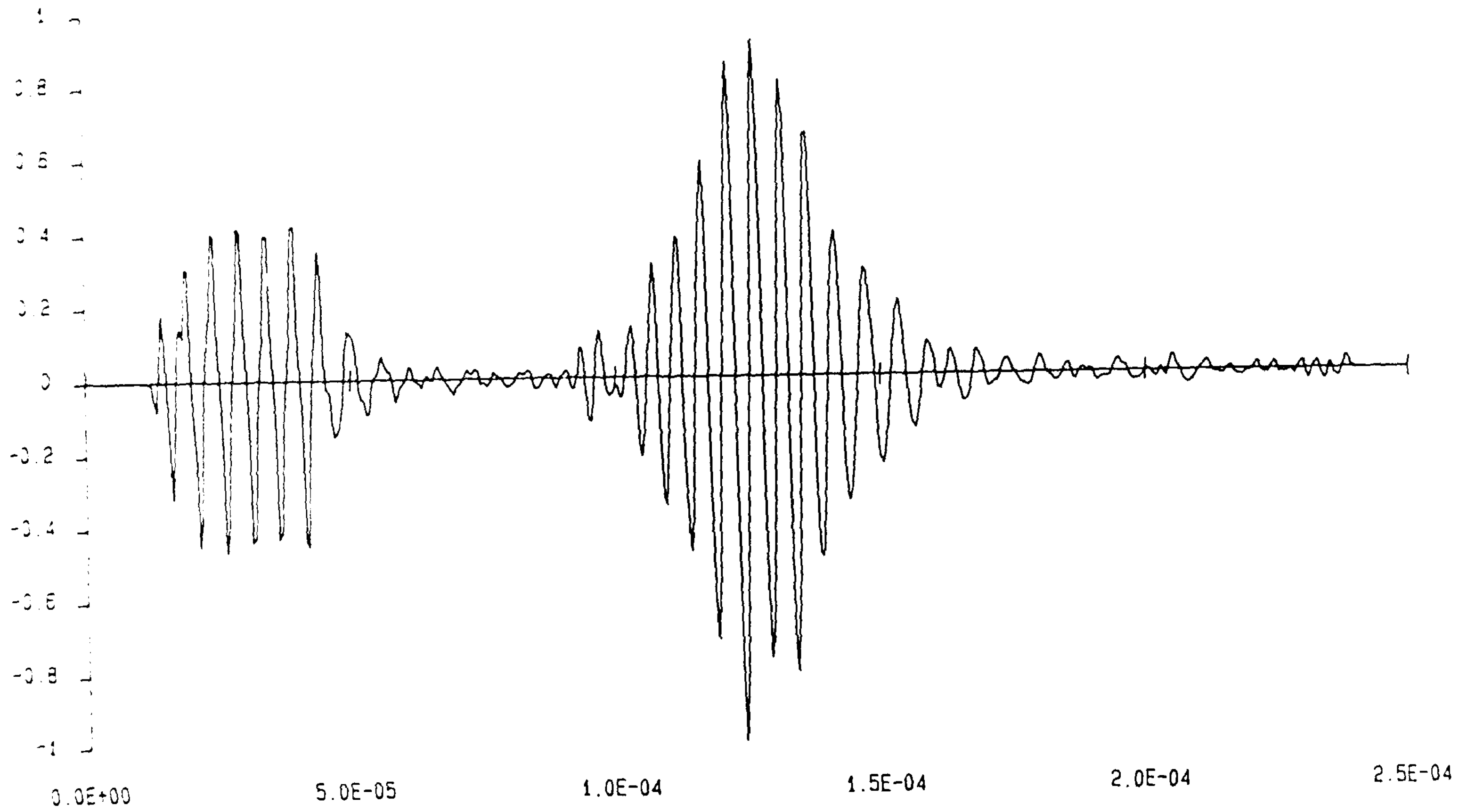

Solid Filled, 5% Shell, 0 Degrees
Time (seconds)
0.0E+00
5.0E-05
1.0E-04
1.5E-04
2.0E-04
2.5E-04
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

ARD

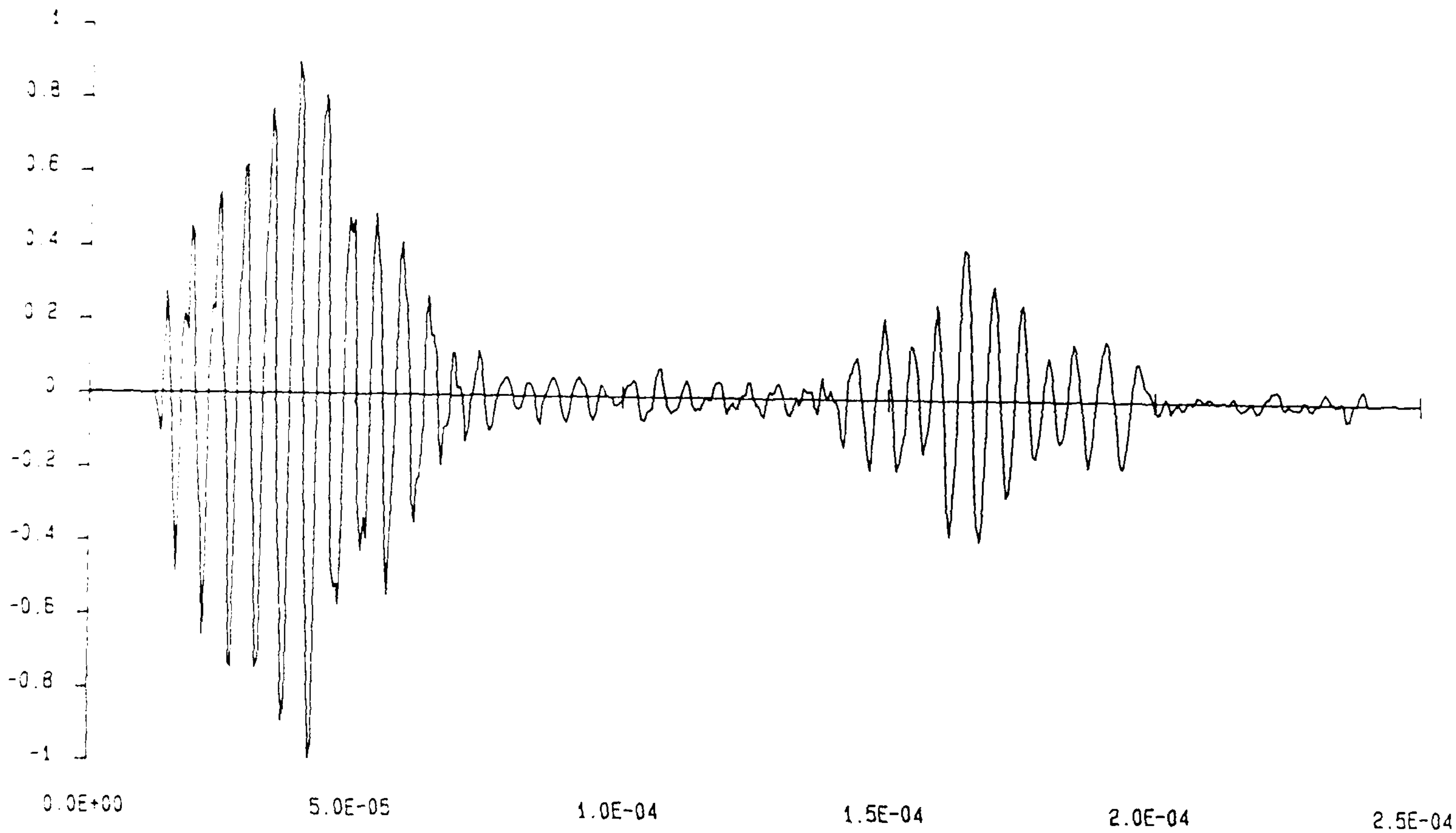

Solid Filled, 10% Shell, 90 Degrees
Time (seconds)
0.0E+00
5.0E-05
1.0E-04
1.5E-04
2.0E-04
2.5E-04
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

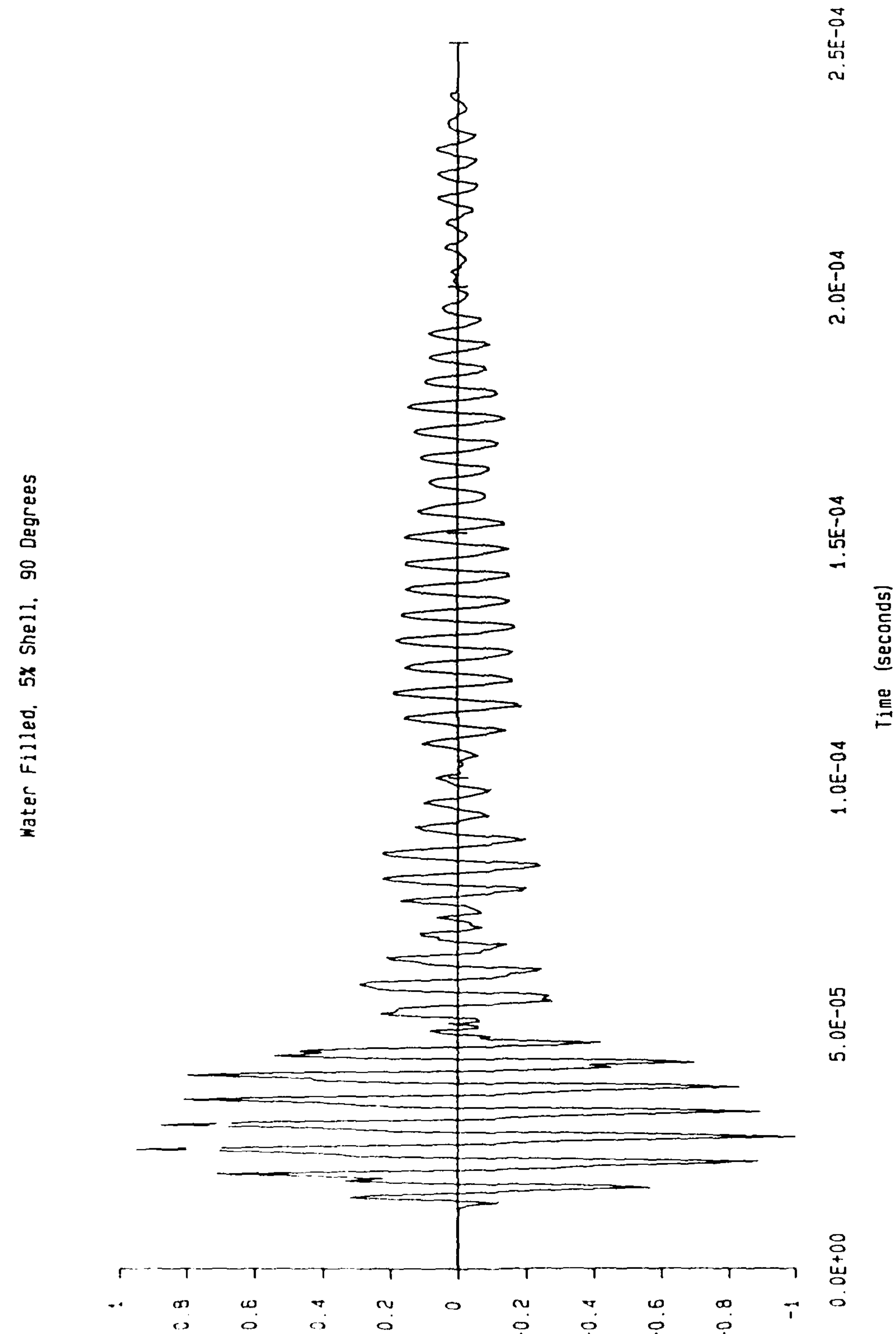
ARD

Solid Filled, 10% Shell, 45 Degrees



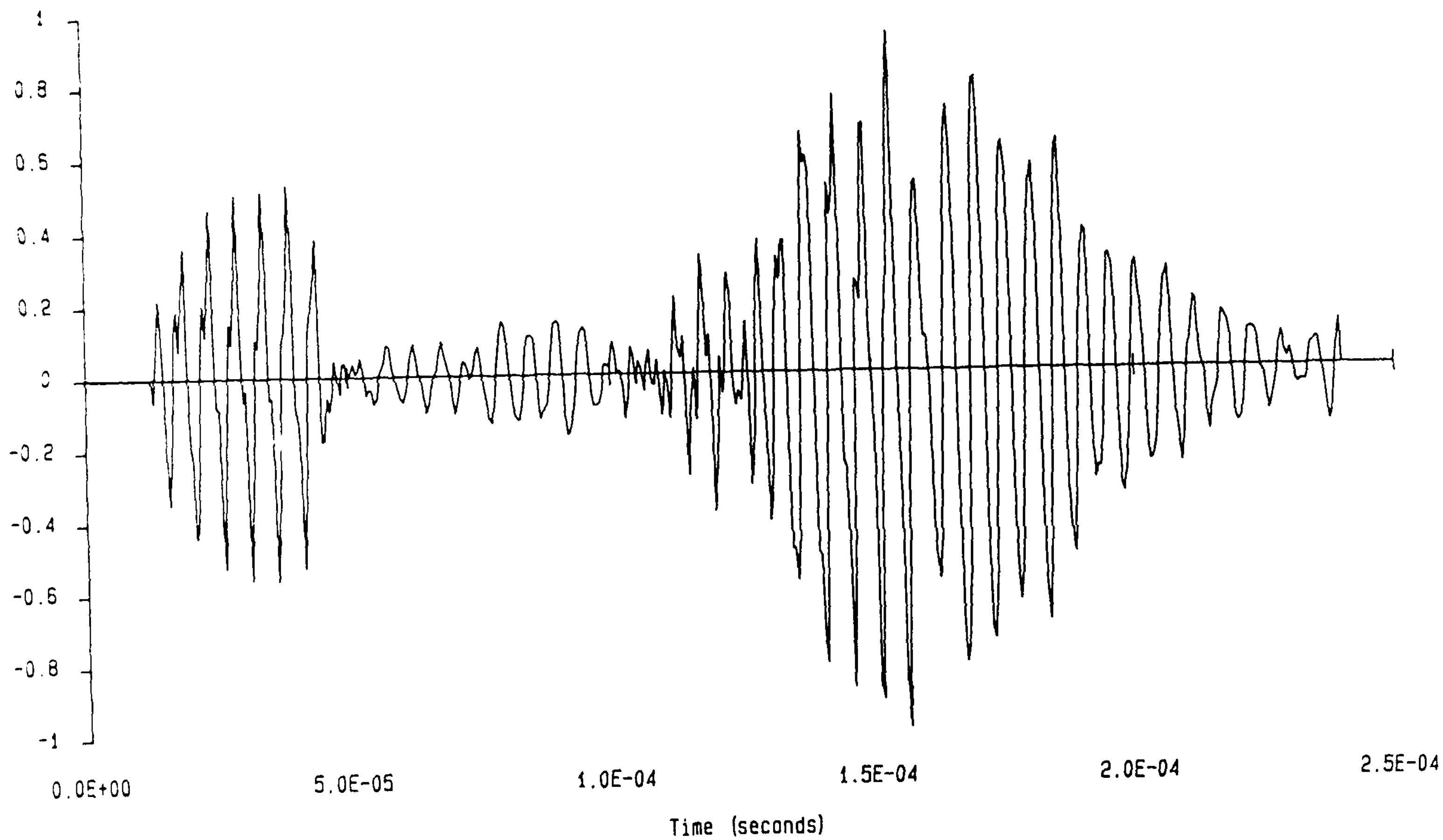
Solid Filled, 10% Shell, 0 Degrees

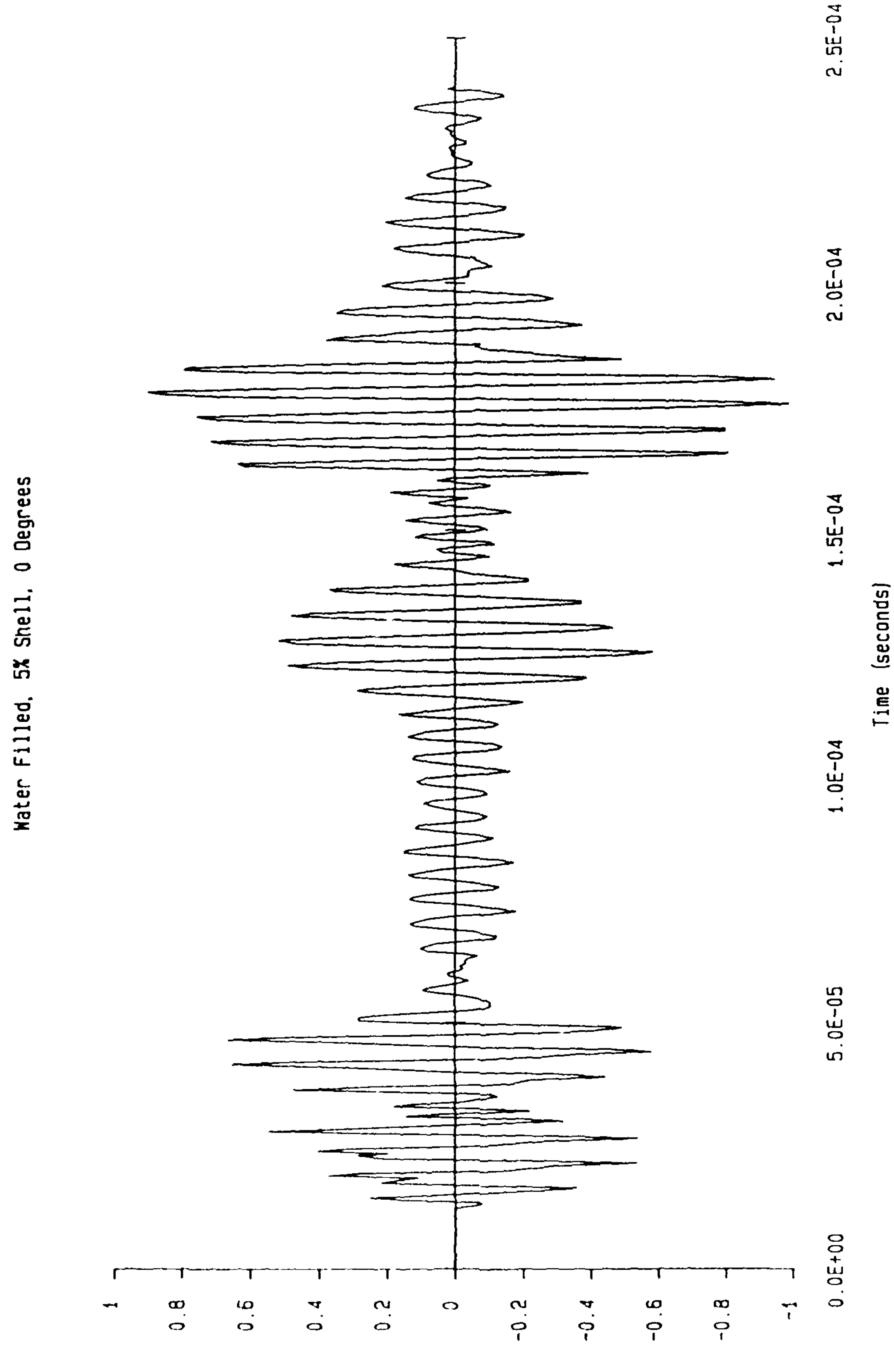



Water Filled, 5% Shell, 90 Degrees
Time (seconds)
0.0E+00
5.0E-05
1.0E-04
1.5E-04
2.0E-04
2.5E-04
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

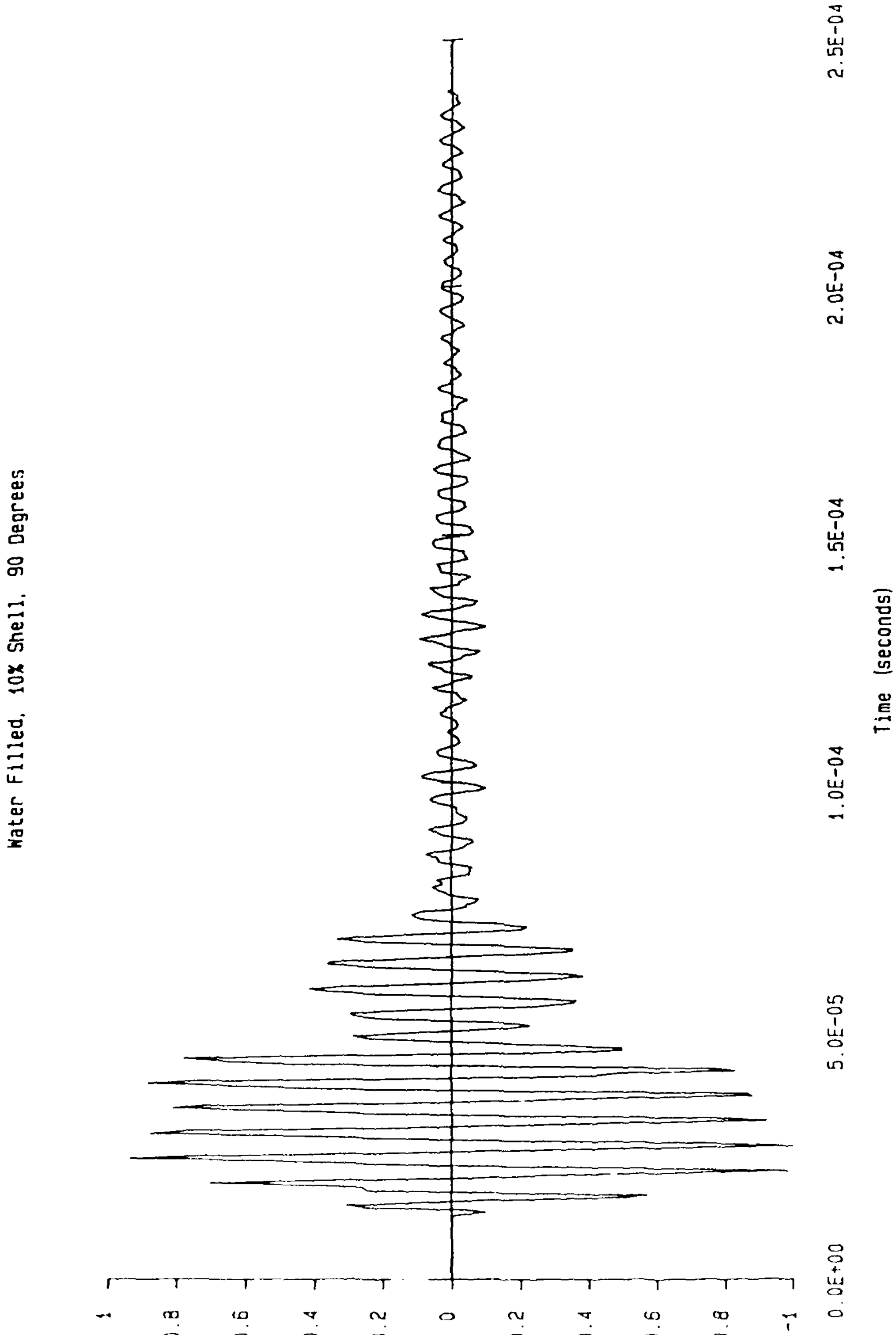
ARD

Water Filled, 5% Shell, 45 Degrees

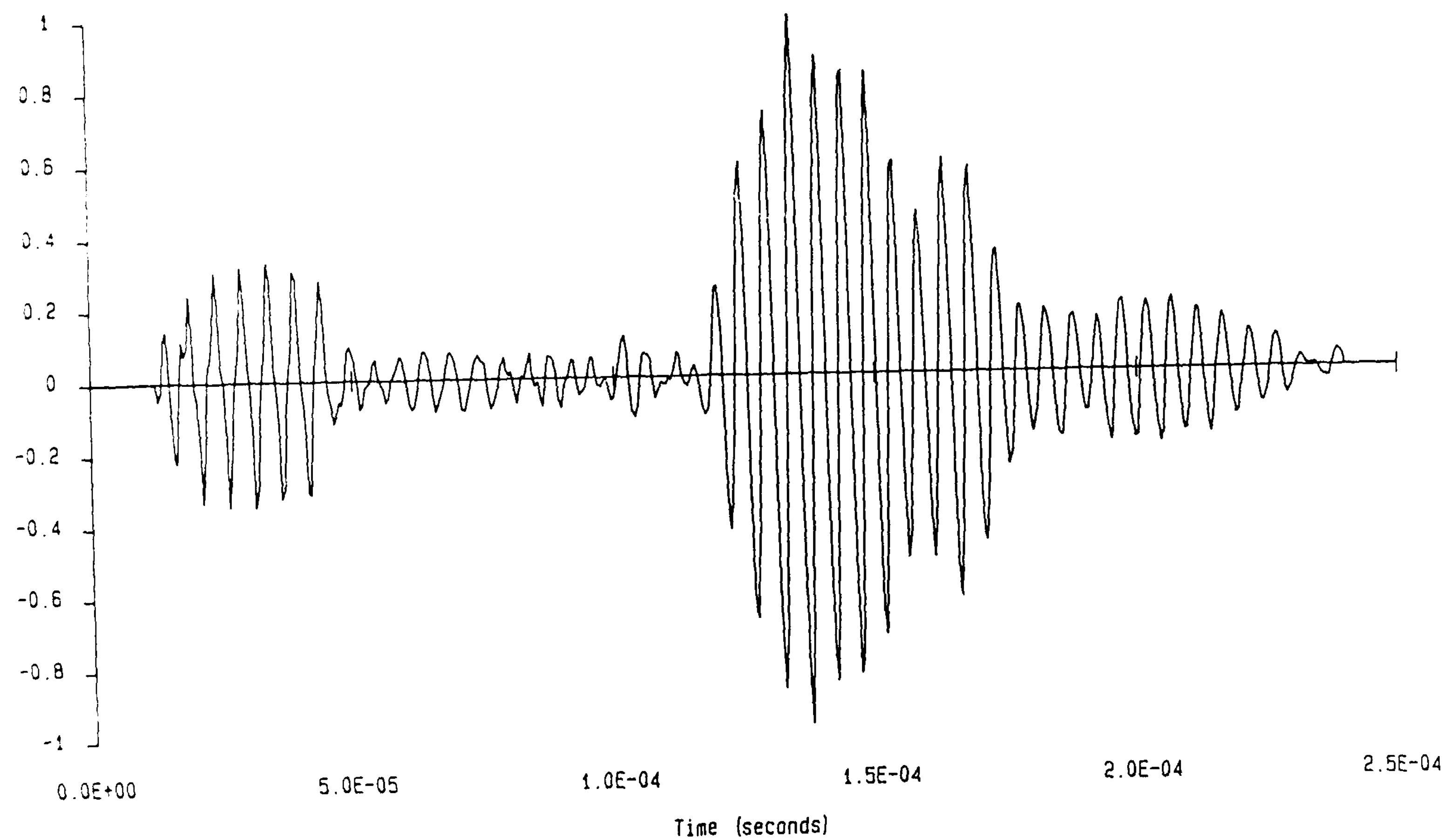




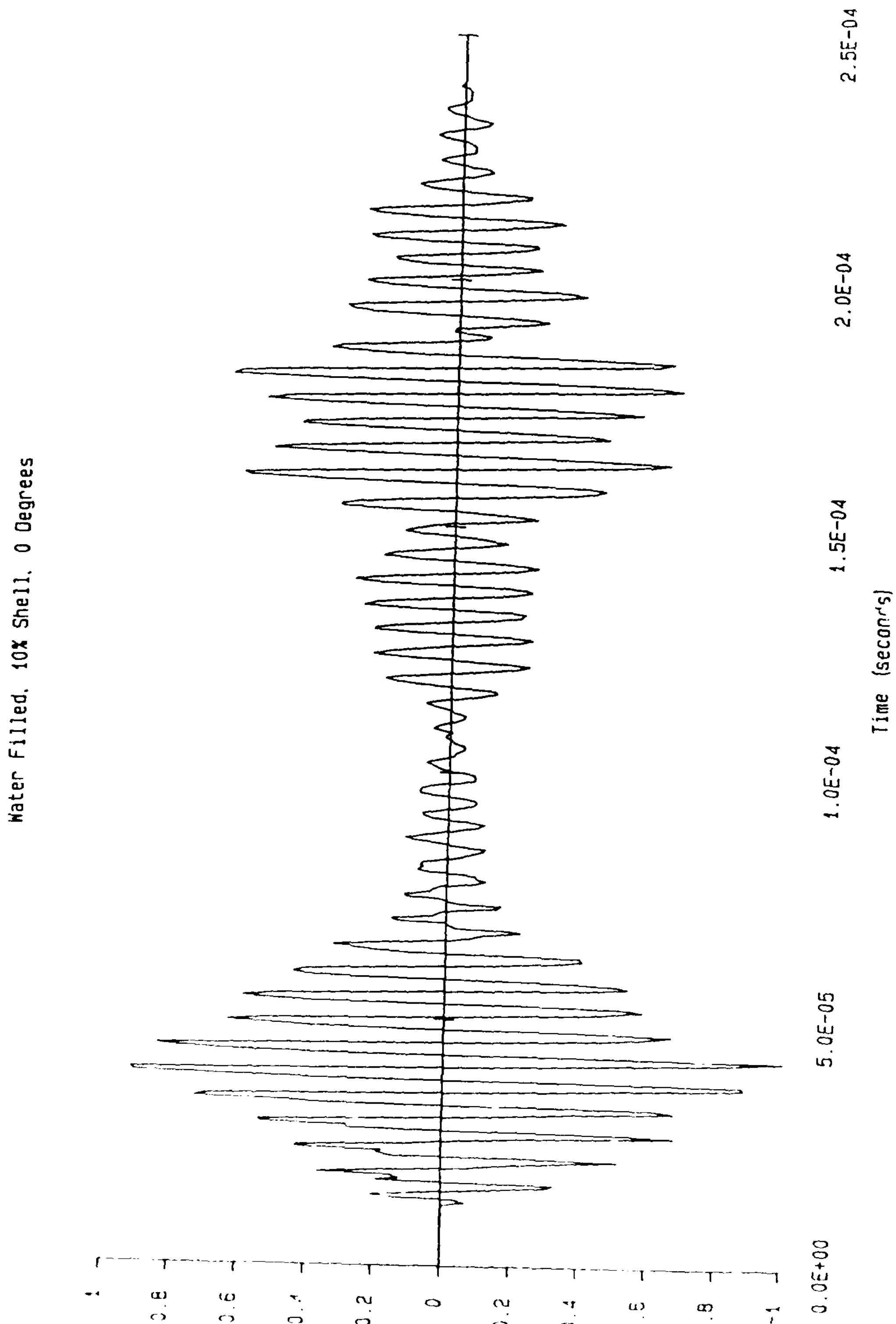
Water Filled, 5% Shell, 0 Degrees
Time (seconds)
0.0E+00
5.0E-05
1.0E-04
1.5E-04
2.0E-04
2.5E-04
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1


Water Filled, 10% Shell, 90 Degrees
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0.0E+00
5.0E-05
1.0E-04
1.5E-04
2.0E-04
2.5E-04
Time (seconds)

Water Filled, 10% Shell, 45 Degrees



ARD

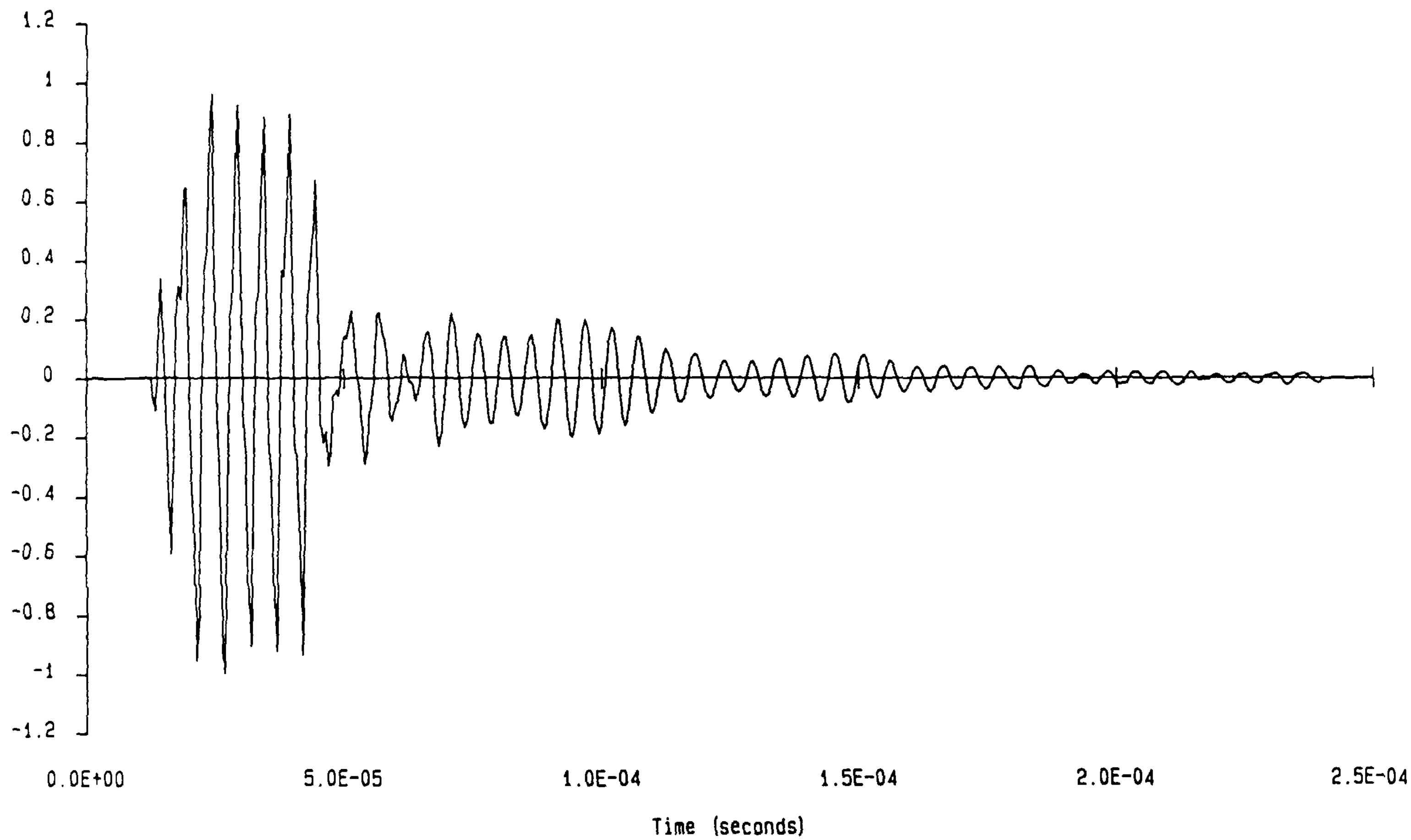

Water Filled, 10% Shell, 0 Degrees
Time (seconds)
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0.0E+00
5.0E-05
1.0E-04
1.5E-04
2.0E-04
2.5E-04

APPENDIX B

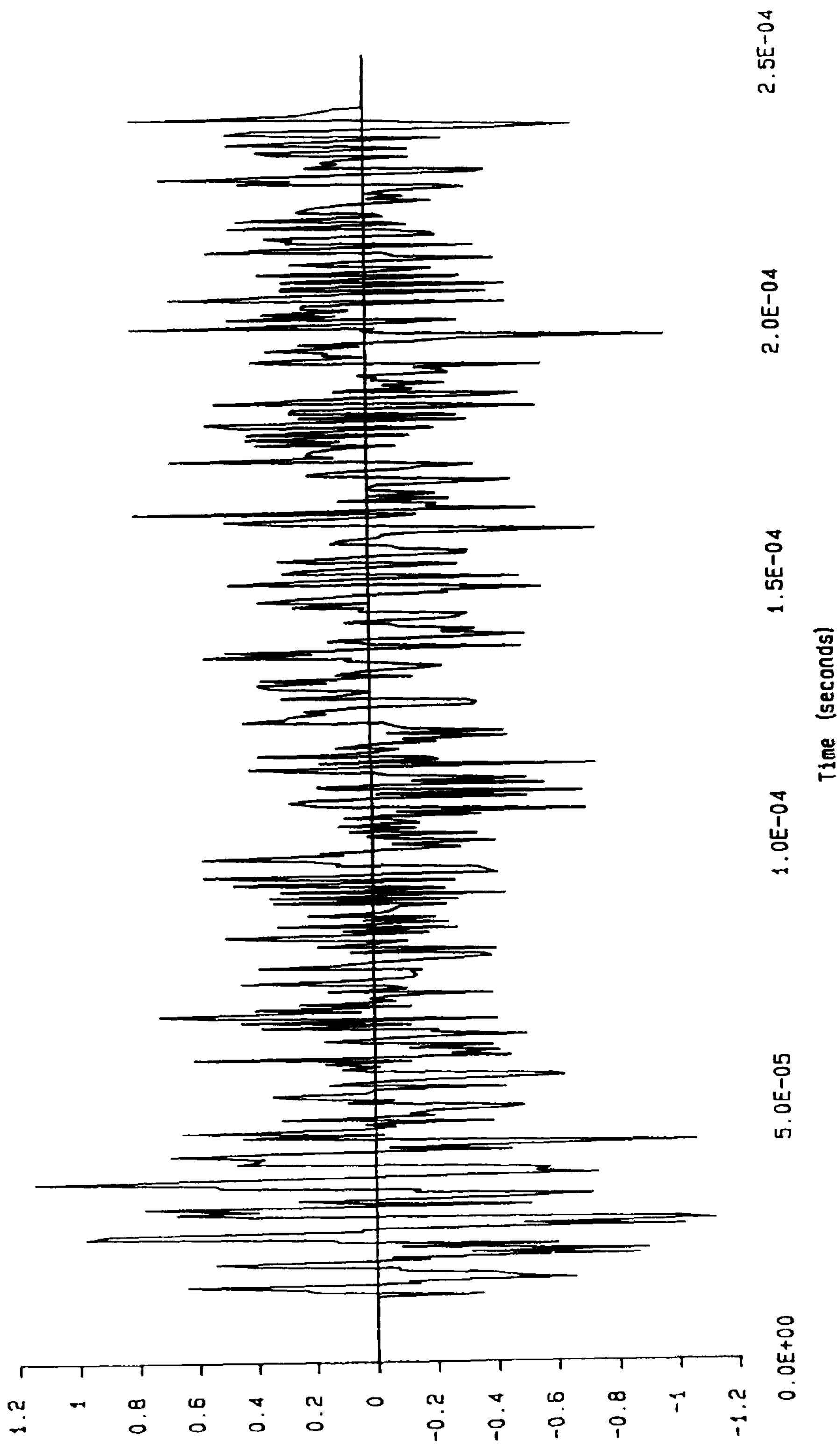
SIGNAL NOISE LEVELS

The airfilled, 5% shell thickness, 90 degree class of signal is shown at three noise levels. The noise levels are clean (the averaged signal), 8.5 dB, and -3.5 dB.

Air Filled, 5% Shell, 90 Degrees, Averaged

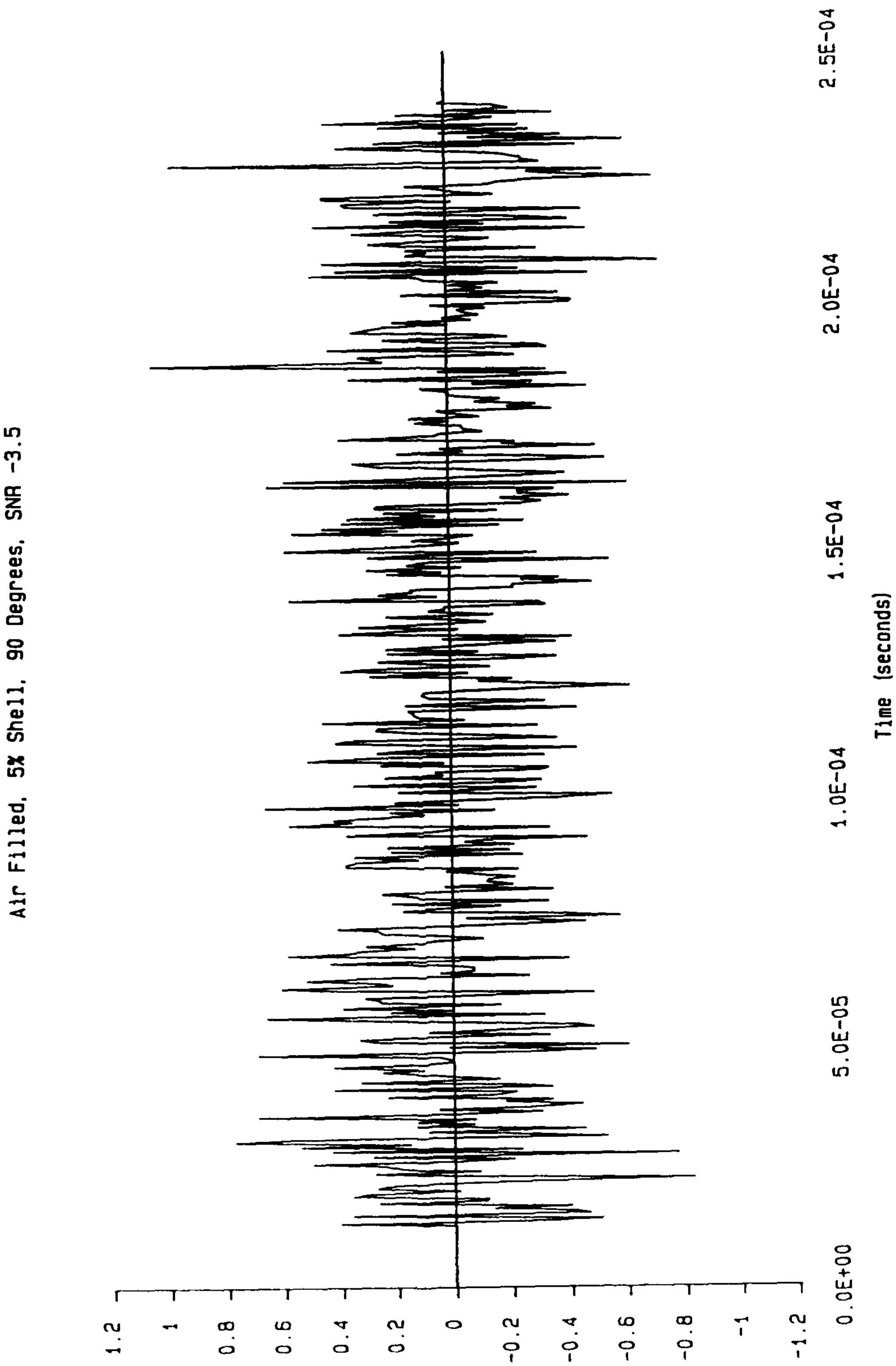


ARD



Air Filled, 5% Shell, 90 Degrees, SNR 8.5
Time (seconds)
0.0E+00
5.0E-05
1.0E-04
1.5E-04
2.0E-04
2.5E-04
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2

ARD

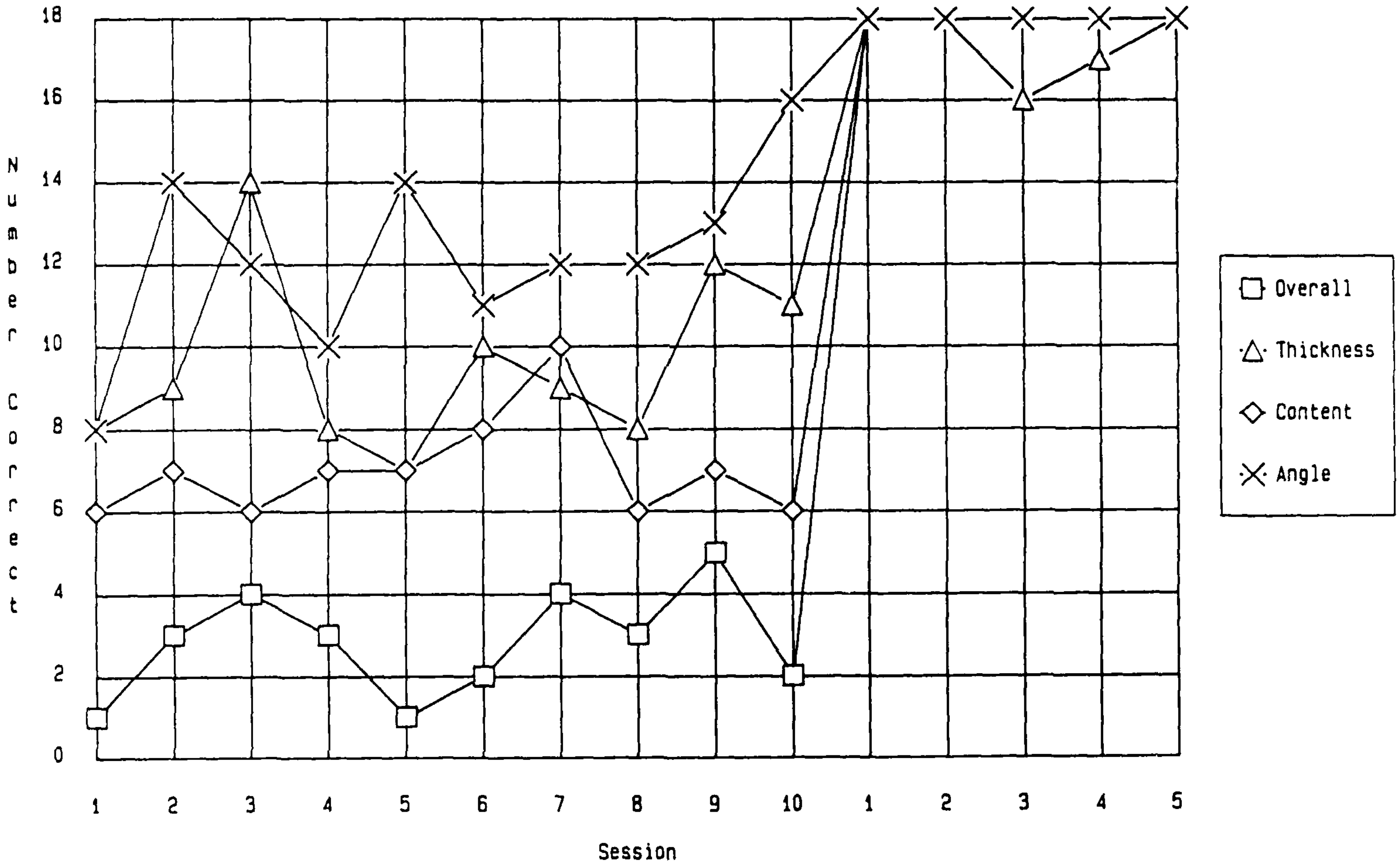

Air Filled, 5% Shell, 90 Degrees, SNR -3.5
Time (seconds)
0.0E+00
5.0E-05
1.0E-04
1.5E-04
2.0E-04
2.5E-04
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2

APPENDIX C

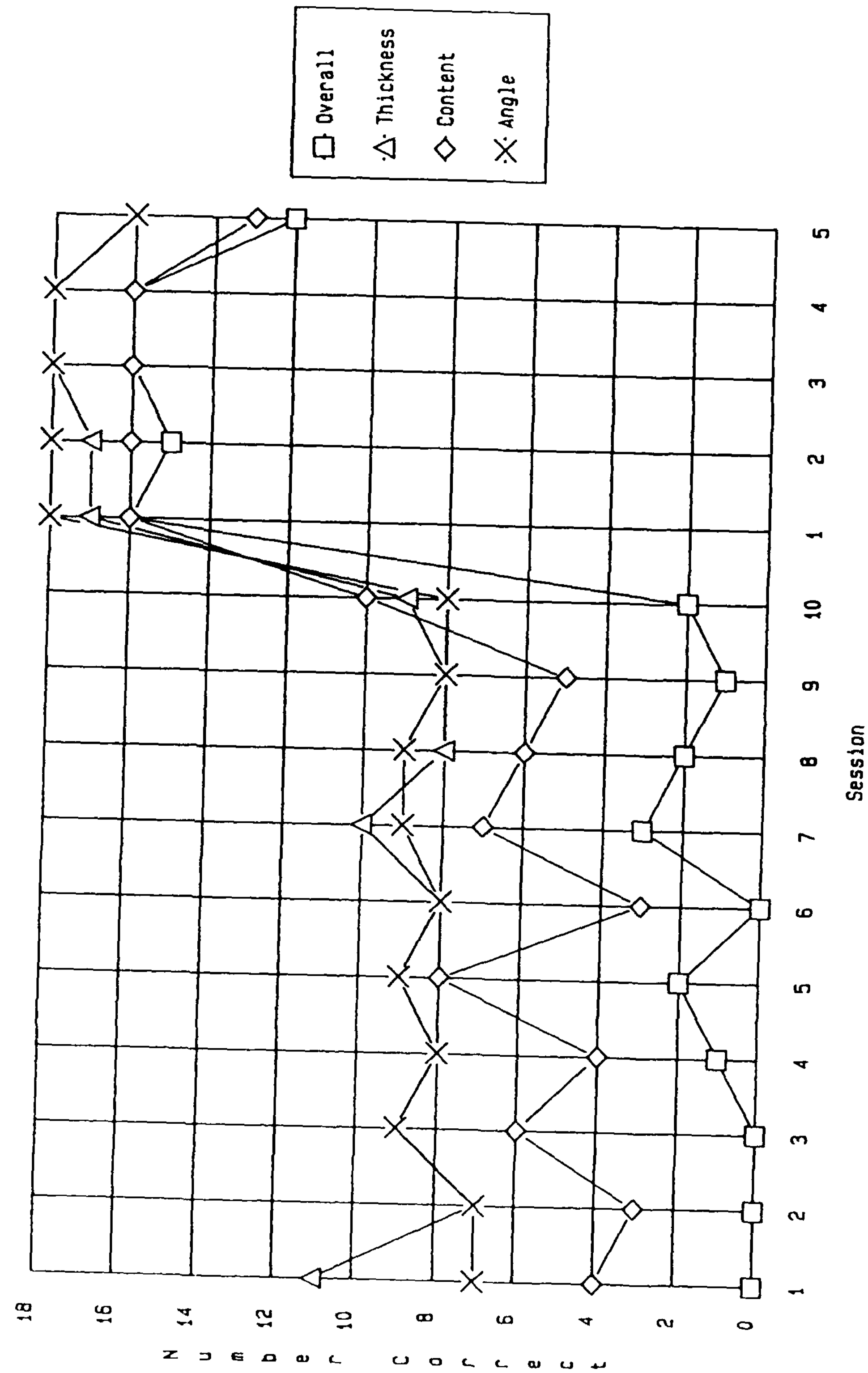
SUBJECT CLASSIFICATION PERFORMANCE

These are plots of the ten subjects' performances. There are two charts for each subject, one for results on clean signals and one for noisy. Each chart has results for the three parameters separately as well as the overall performance. The first ten points on each chart are the results of Experiment 1 and the last five points are results of Experiment 3.

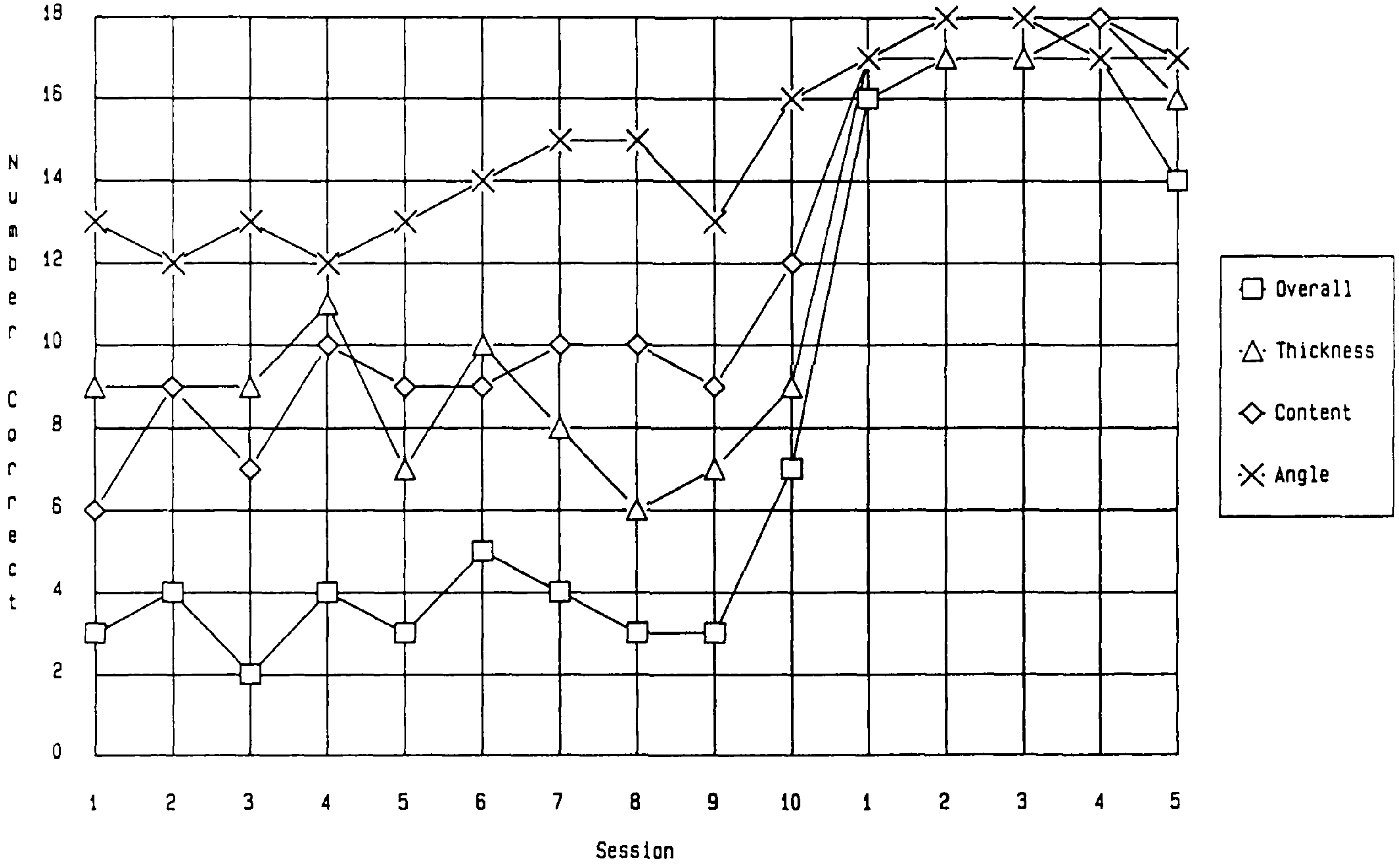
Subject 1 Performance on Clean Signals



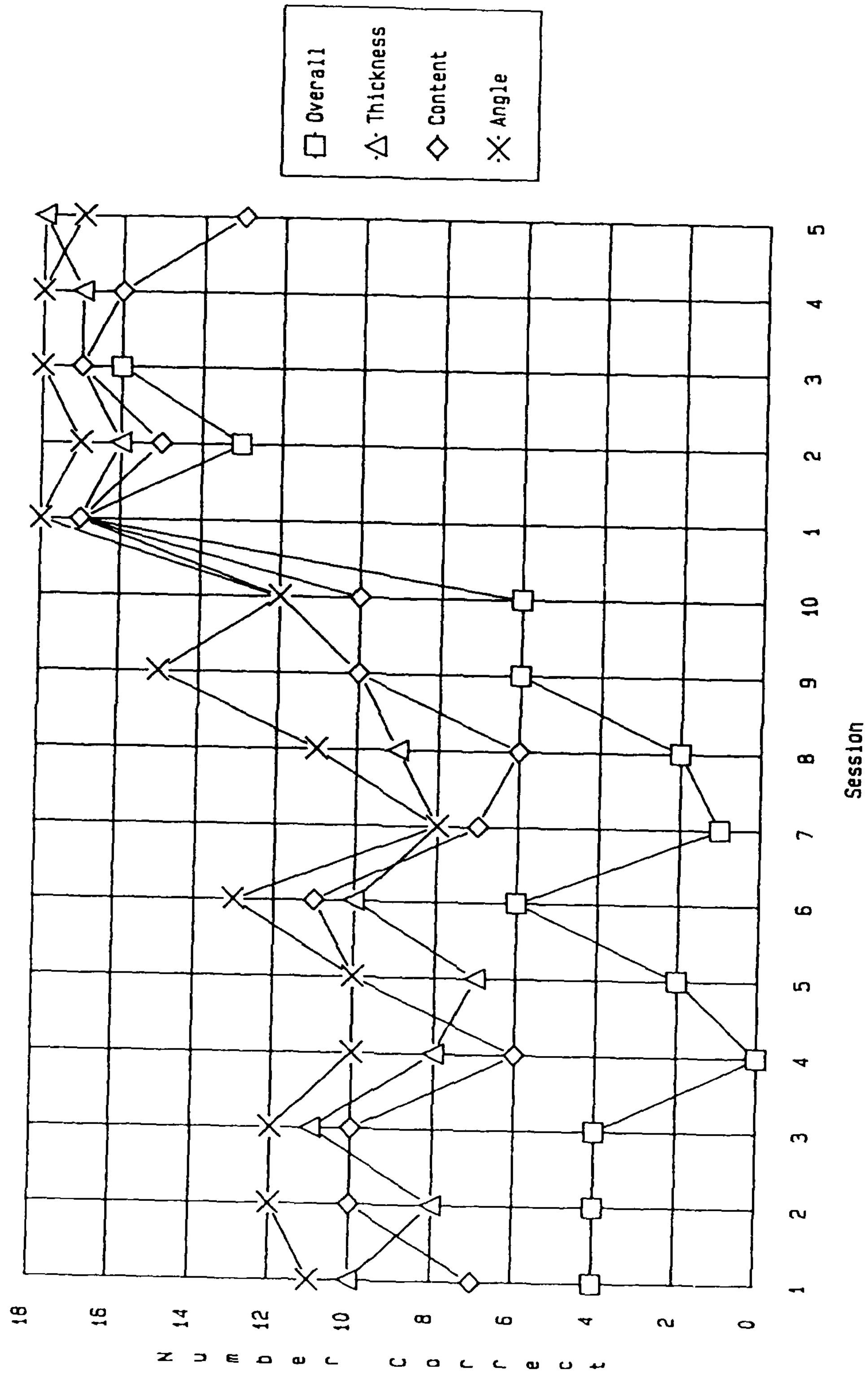
Subject 1 Performance on Noisy Signals



Subject 2 Performance on Clean Signals

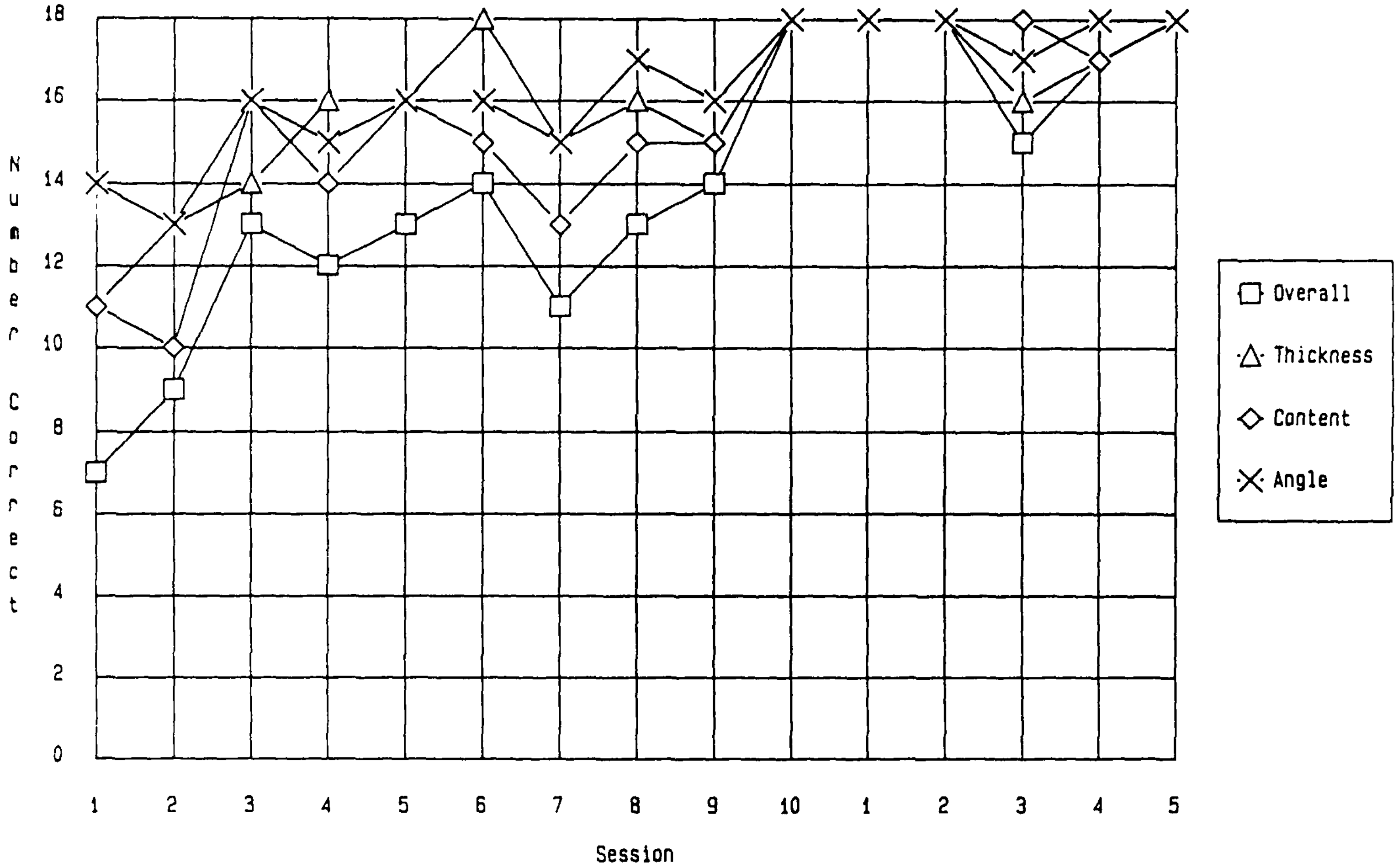


Subject 2 Performance on Noisy Signals

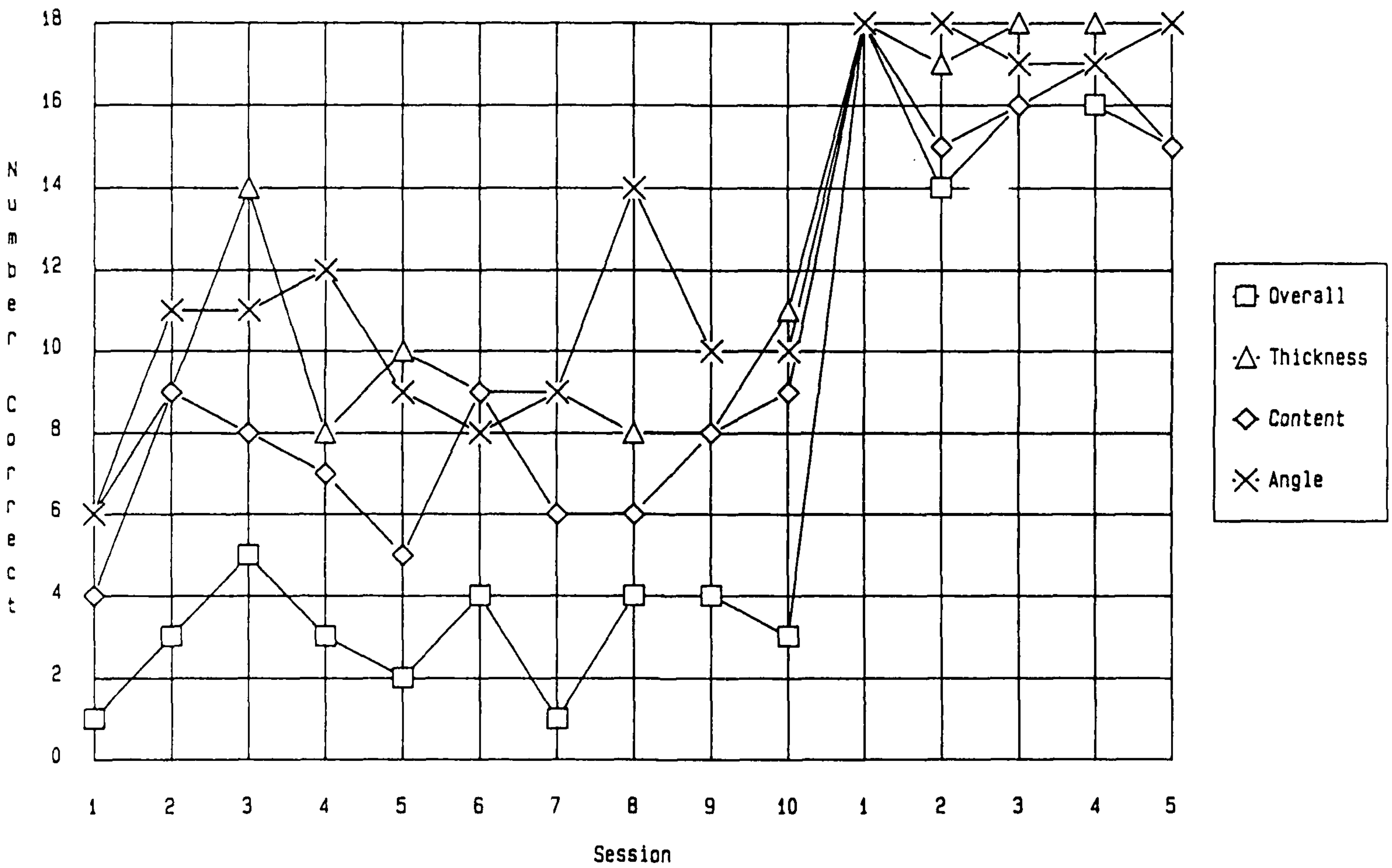


ARD

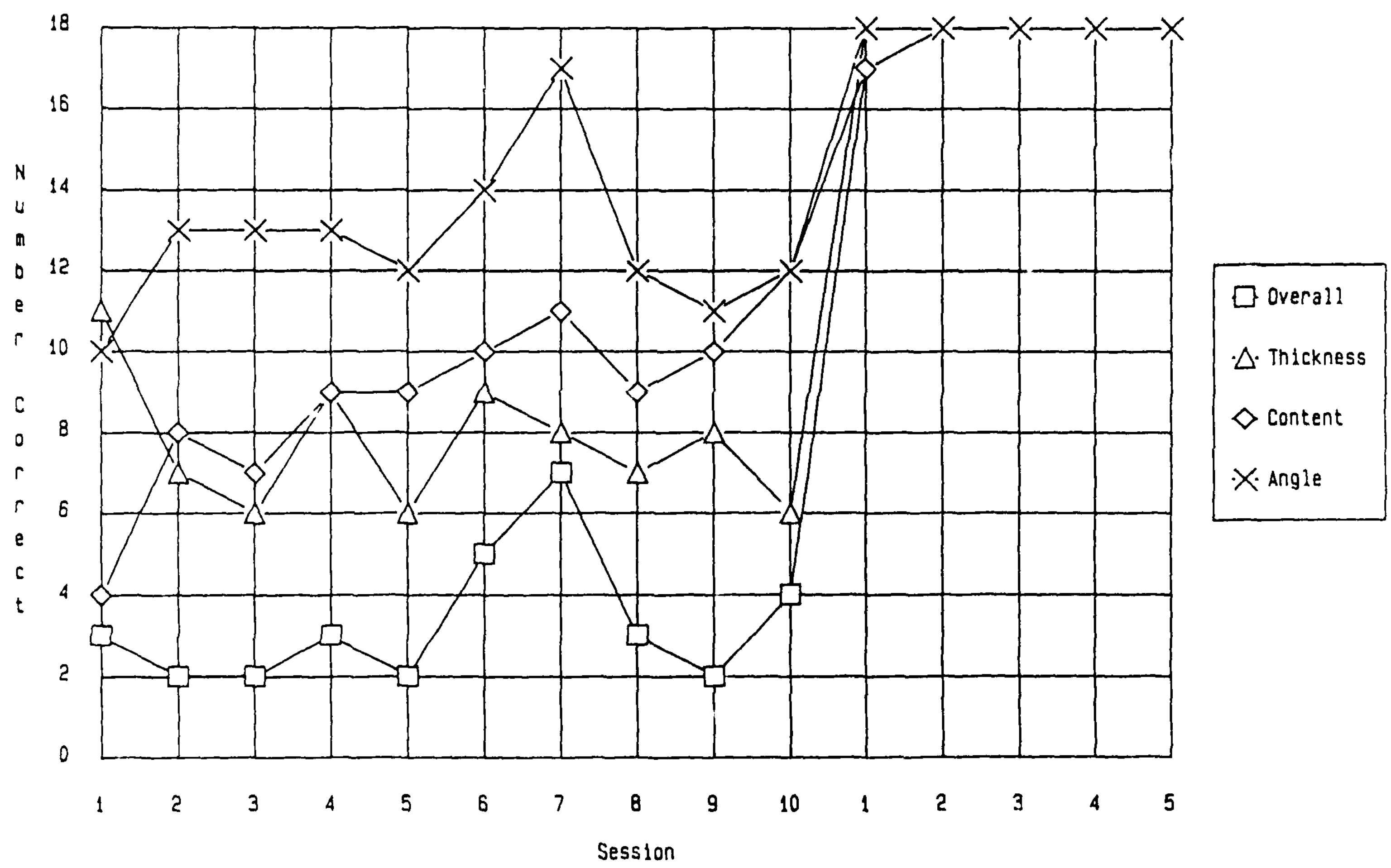
Subject 3 Performance on Clean Signals



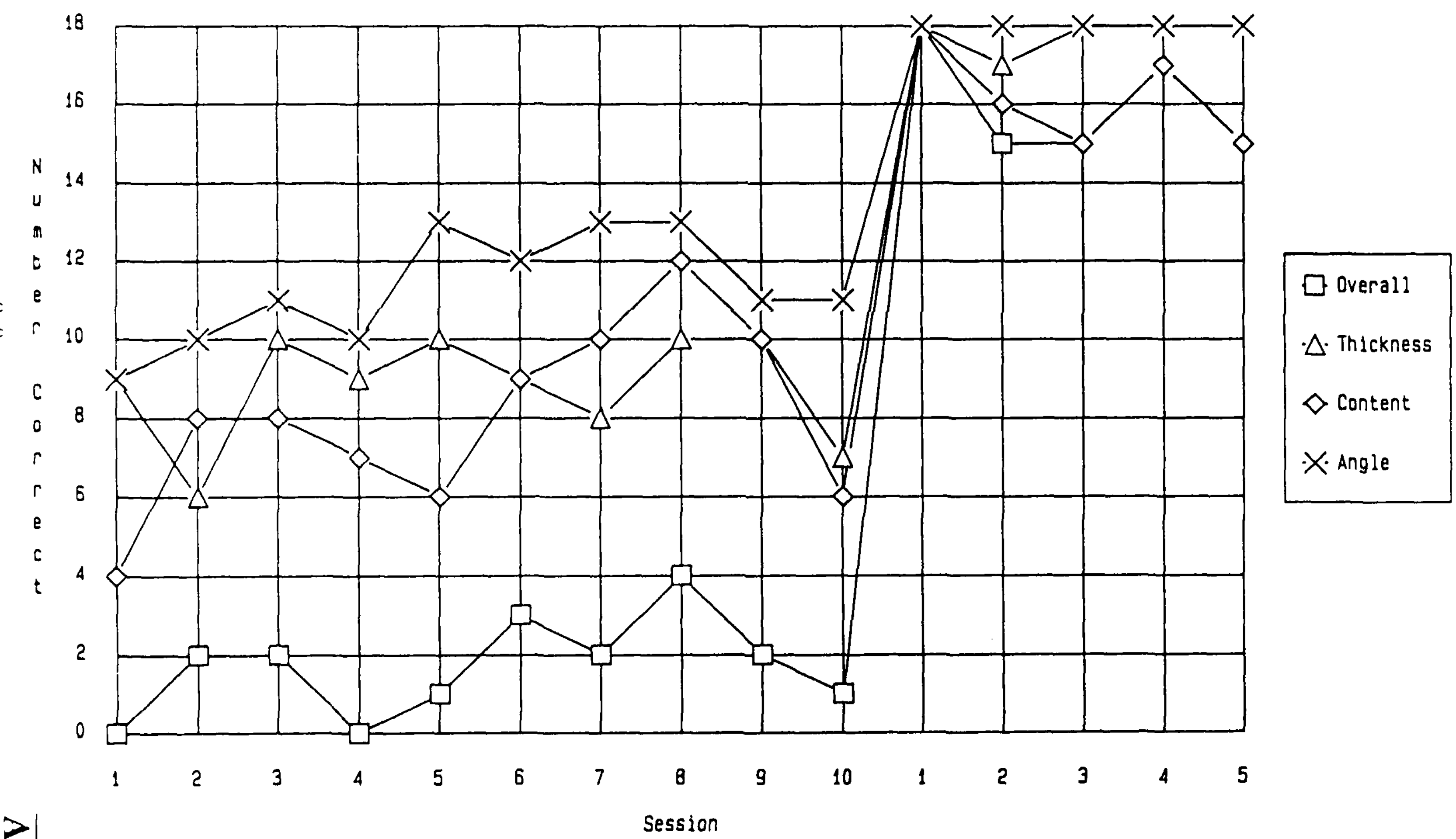
Subject 3 Performance on Noisy Signals



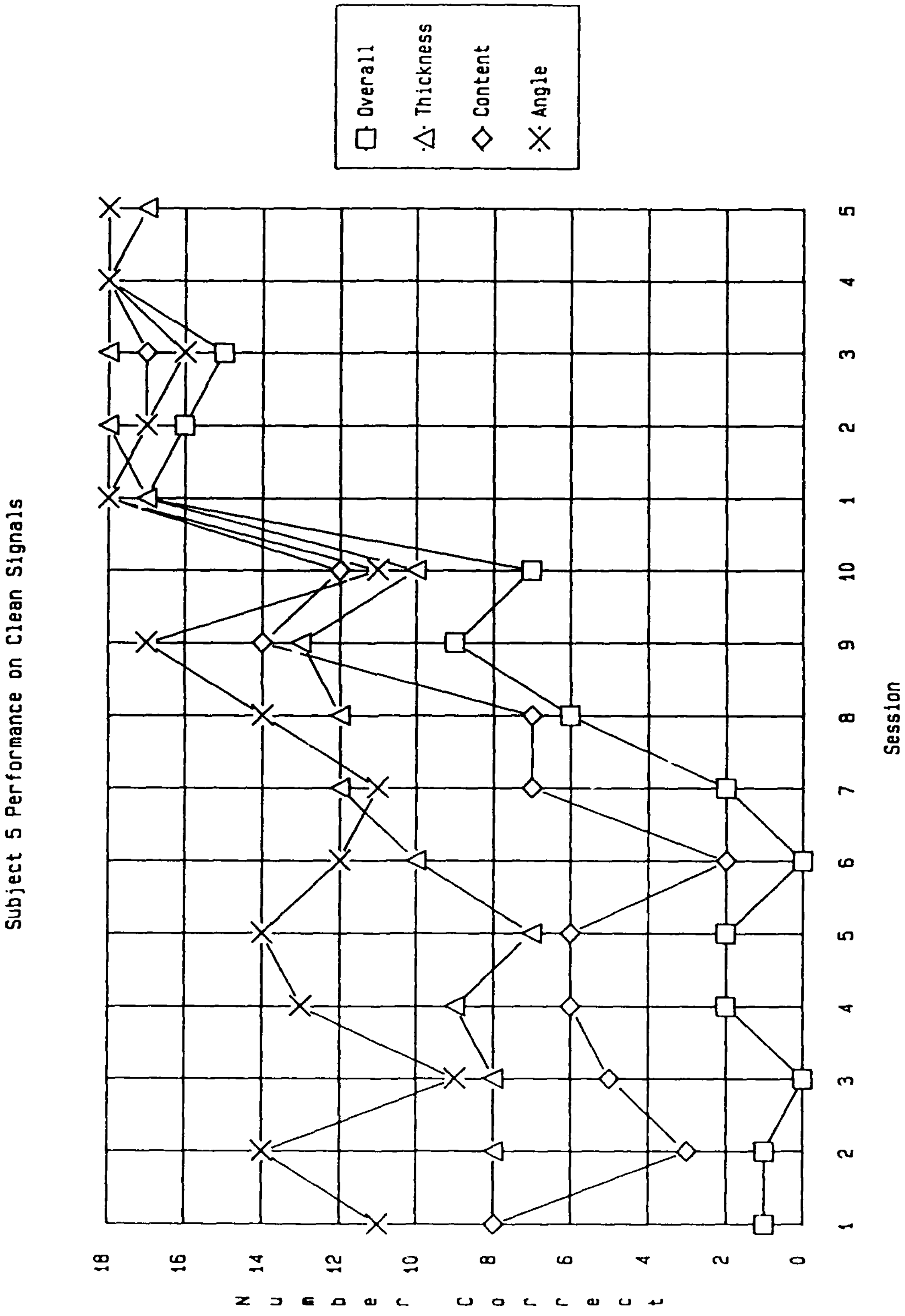
Subject 4 Performance on Clean Signals



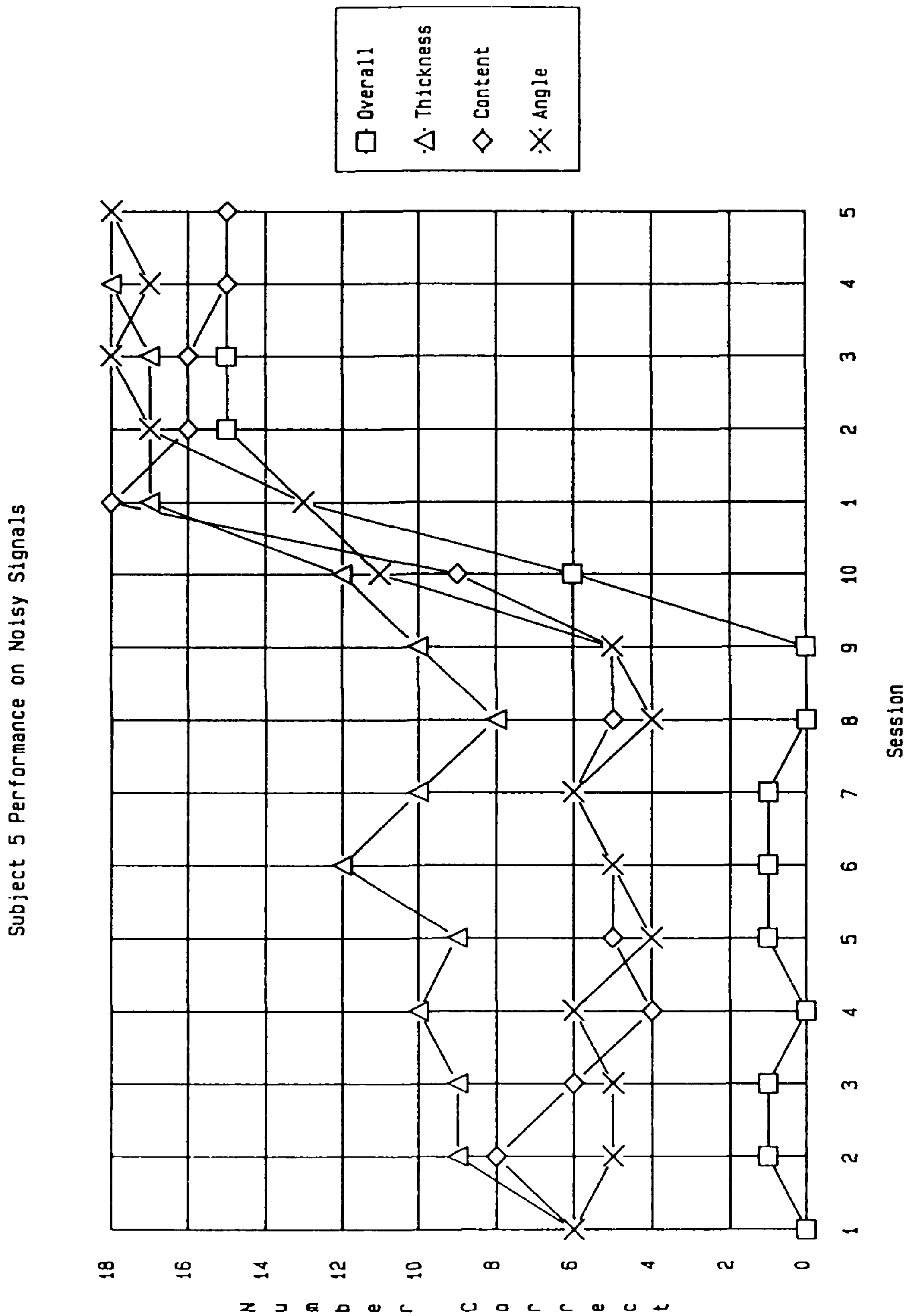
Subject 4 Performance on Noisy Signals

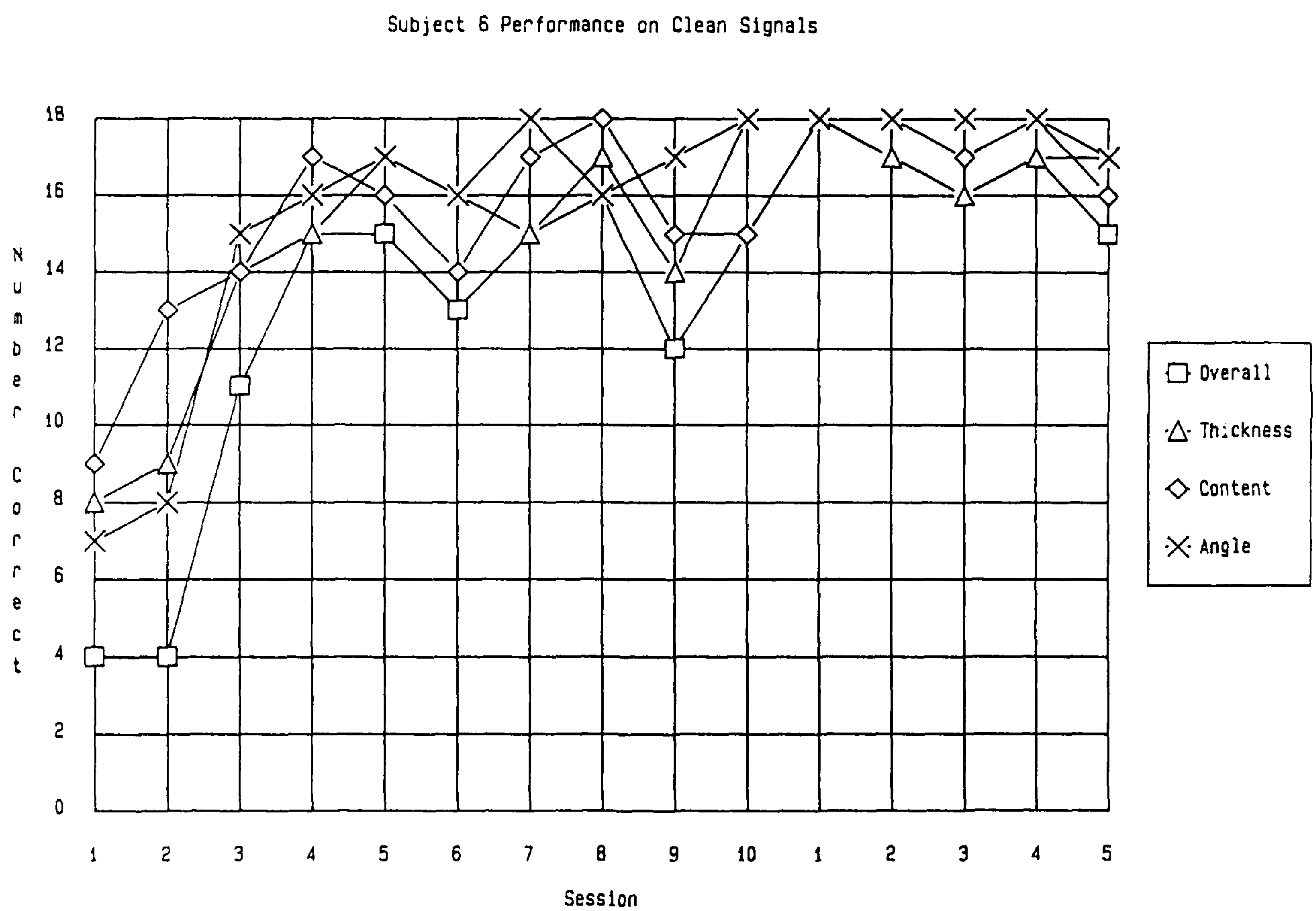


ARD

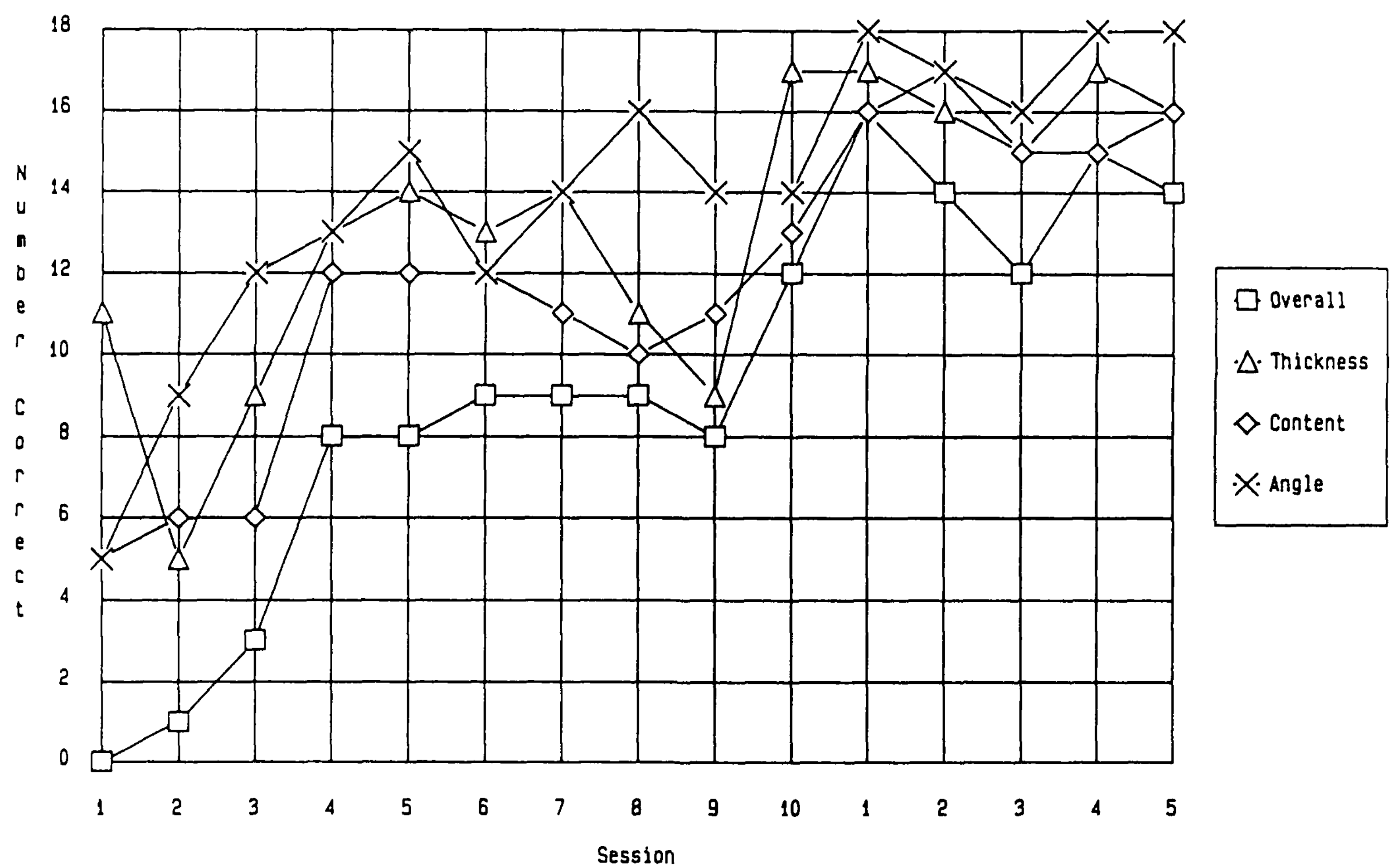

Subject 5 Performance on Clean Signals
Overall
Thickness
Content
Angle
Number Correct
Session
0
2
4
6
8
10
12
14
16
18
1
2
3
4
5
6
7
8
9
10
1
2
3
4
5

ARD


Subject 5 Performance on Noisy Signals
Overall
Thickness
Content
Angle
Number Correct
Session
0
2
4
6
8
10
12
14
16
18
1
2
3
4
5
6
7
8
9
10
1
2
3
4
5

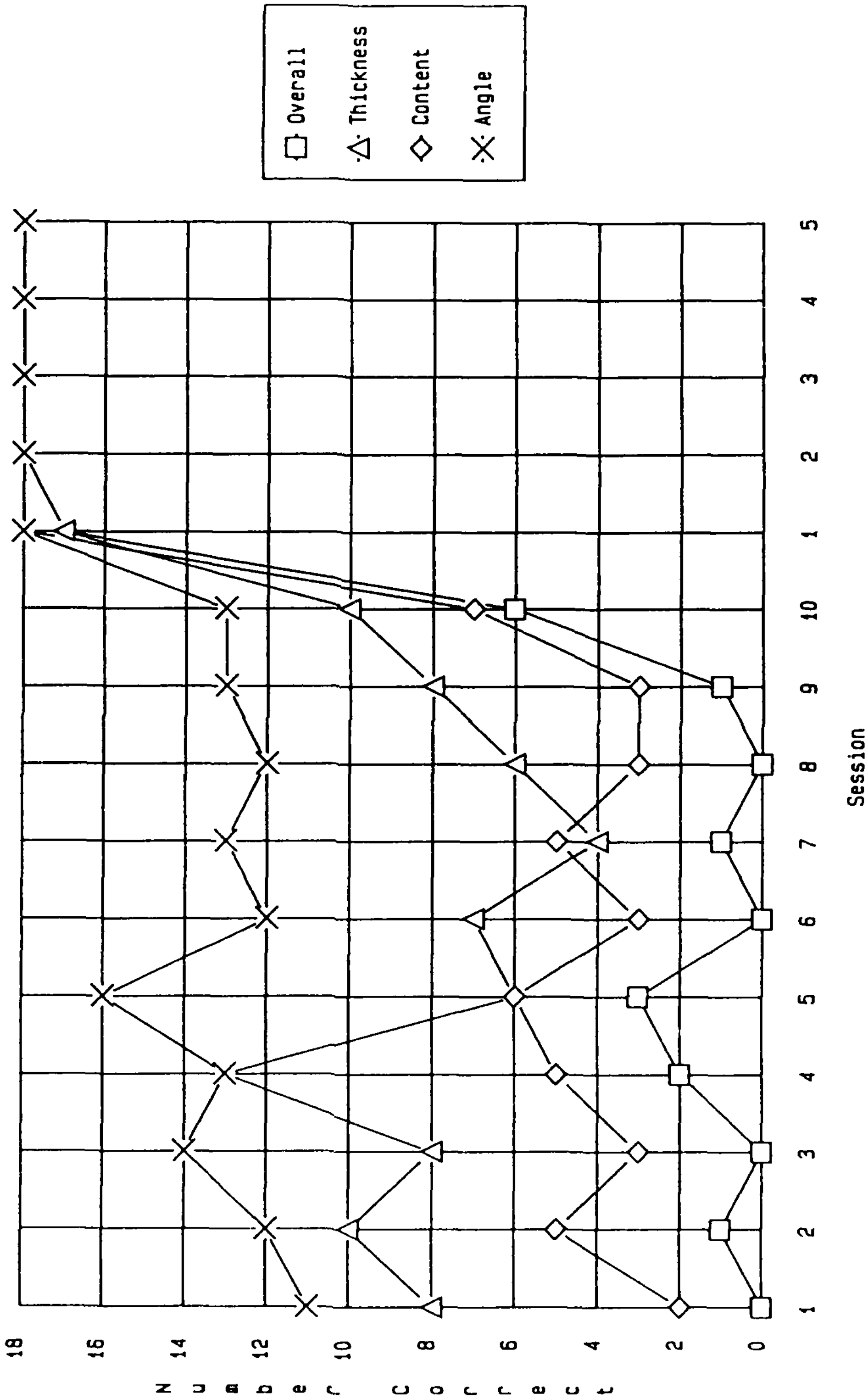

Subject 6 Performance on Clean Signals
Number Correct
Session
Overall
Thickness
Content
Angle

Subject 6 Performance on Noisy Signals

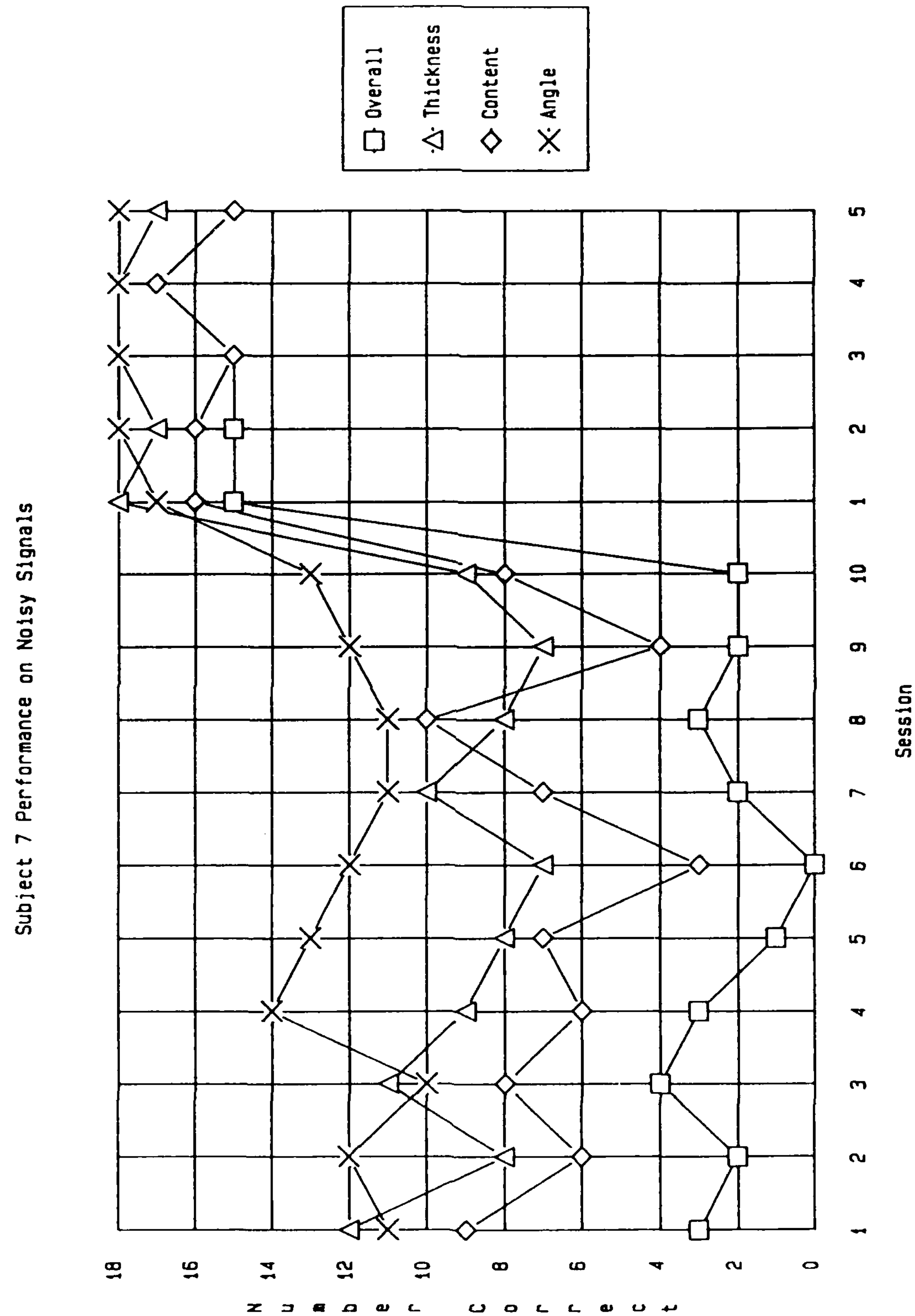


ARD

Subject 7 Performance on Clean Signals

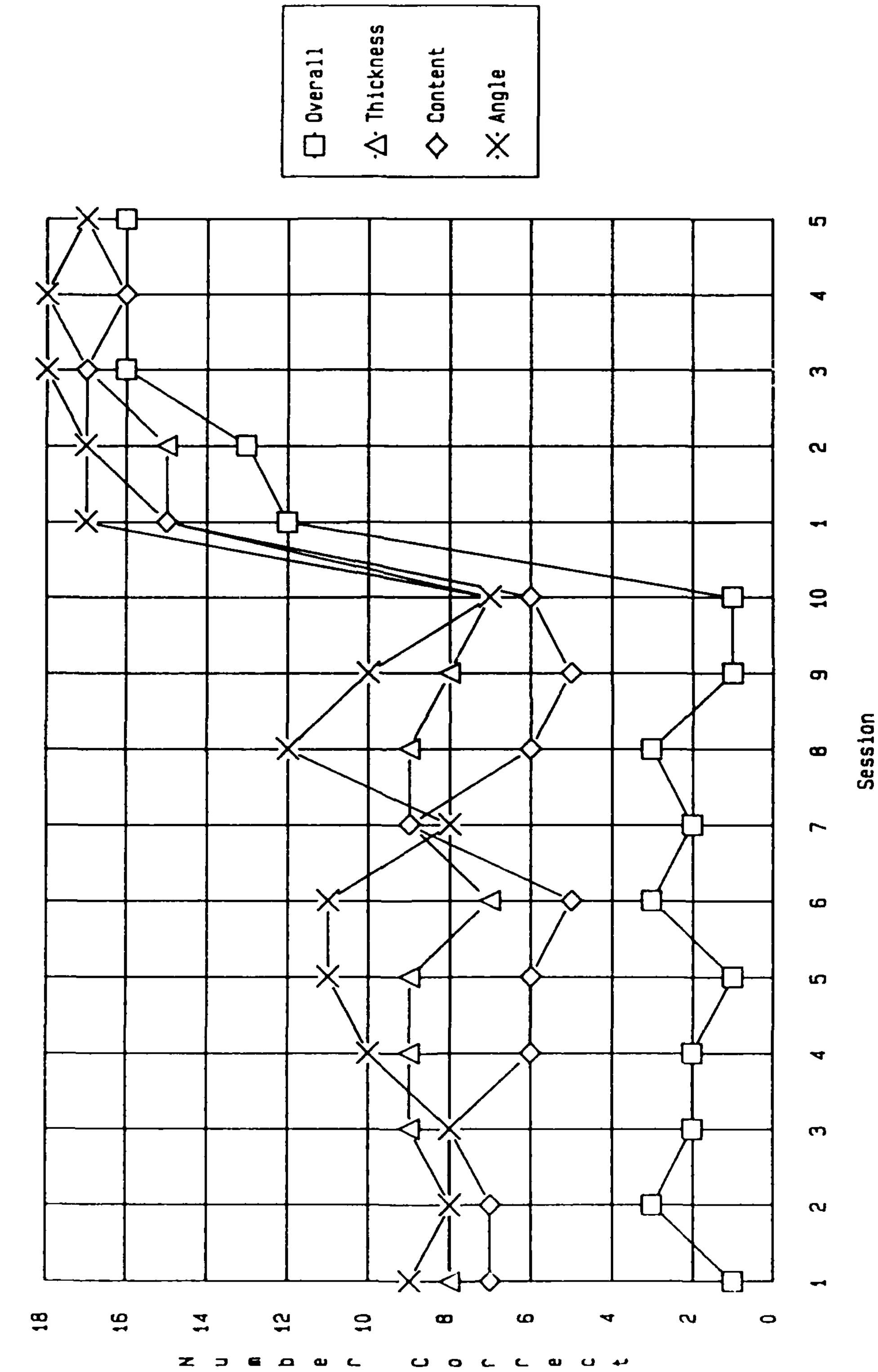


ARD



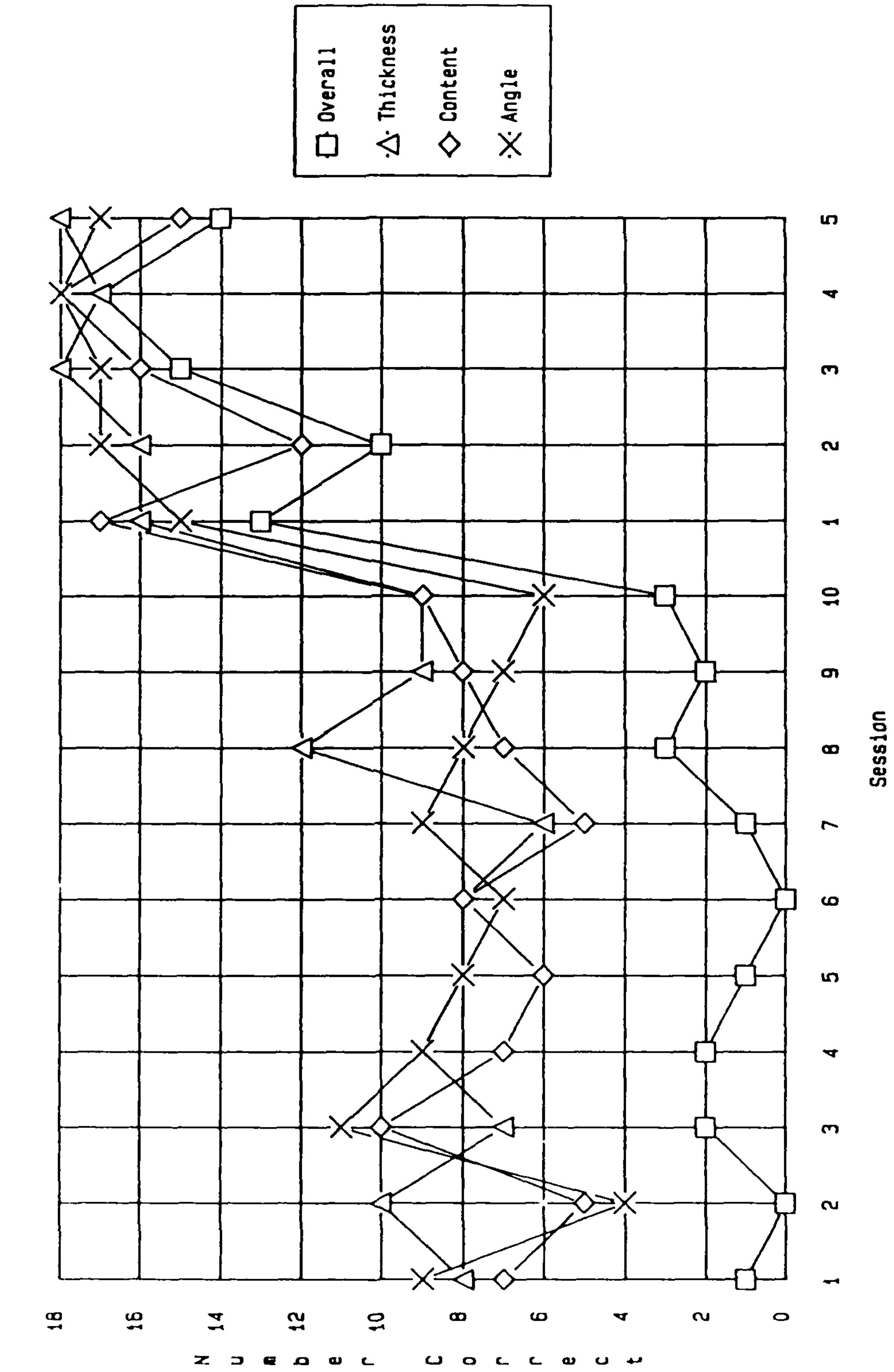
Subject 7 Performance on Noisy Signals
Overall
Thickness
Content
Angle
Number Correct
0
2
4
6
8
10
12
14
16
18
Session
1 2 3 4 5 6 7 8 9 10 1 2 3 4 5

Subject 8 Performance on Clean Signals

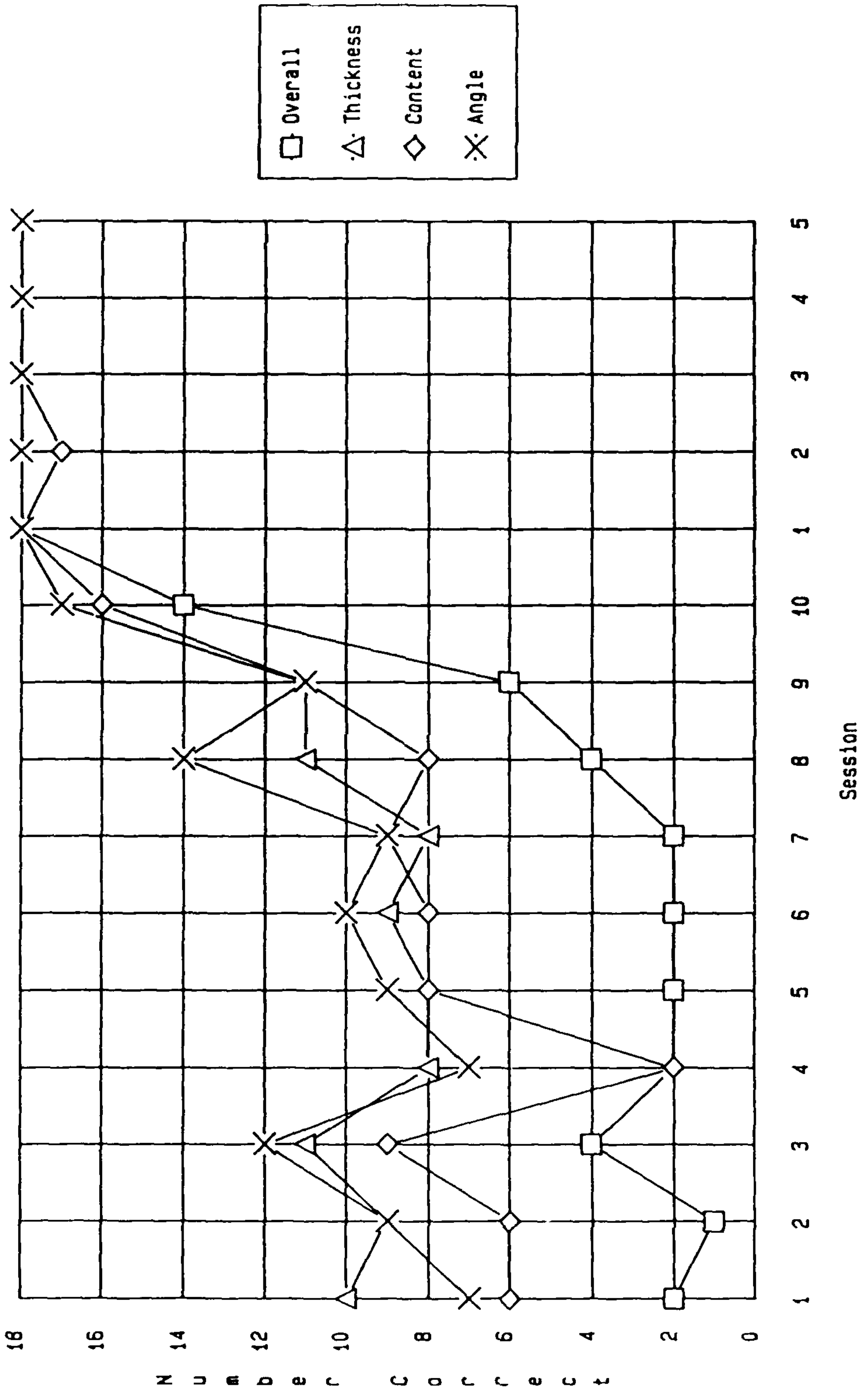


ARD

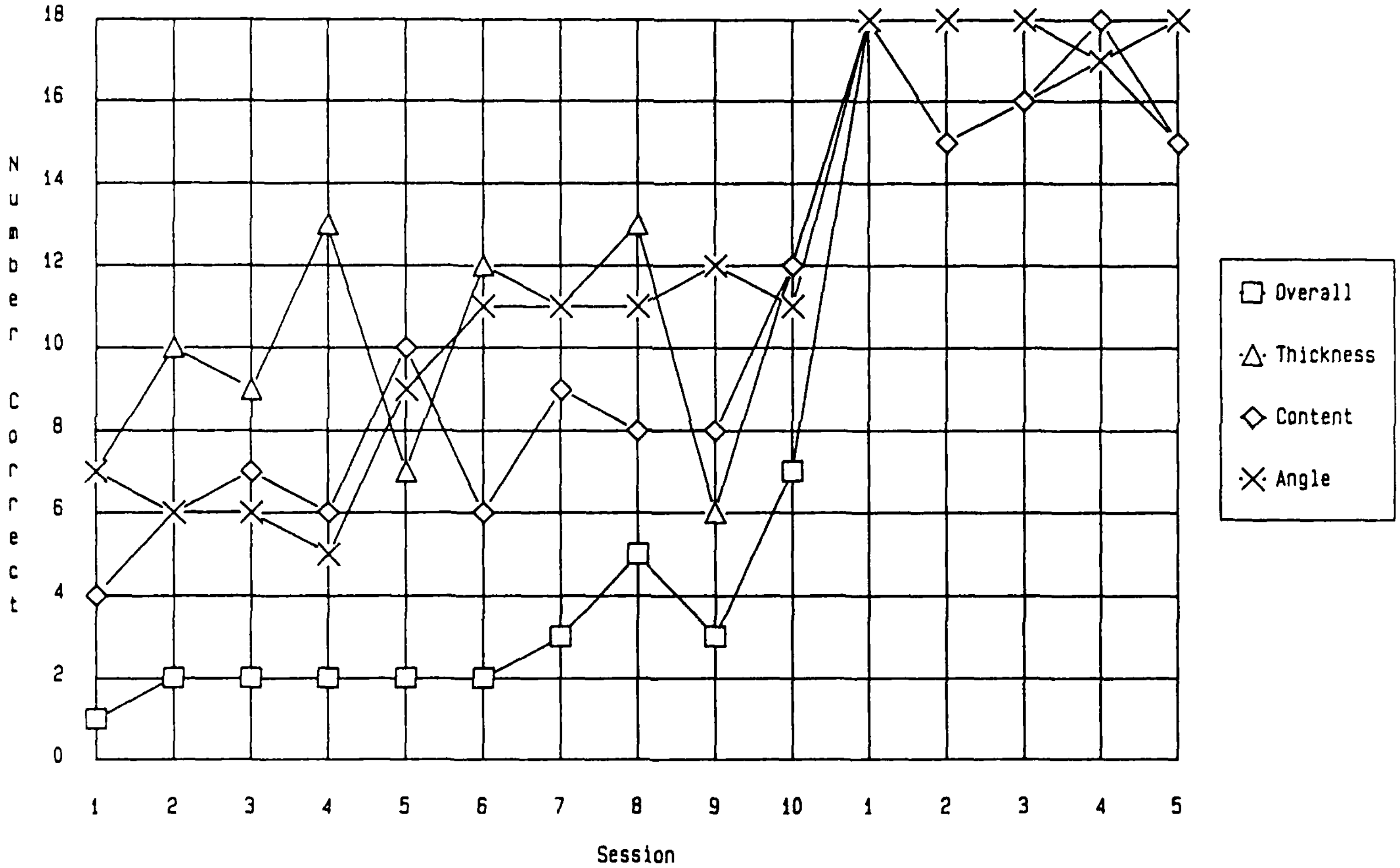
Subject 8 Performance on Noisy Signals

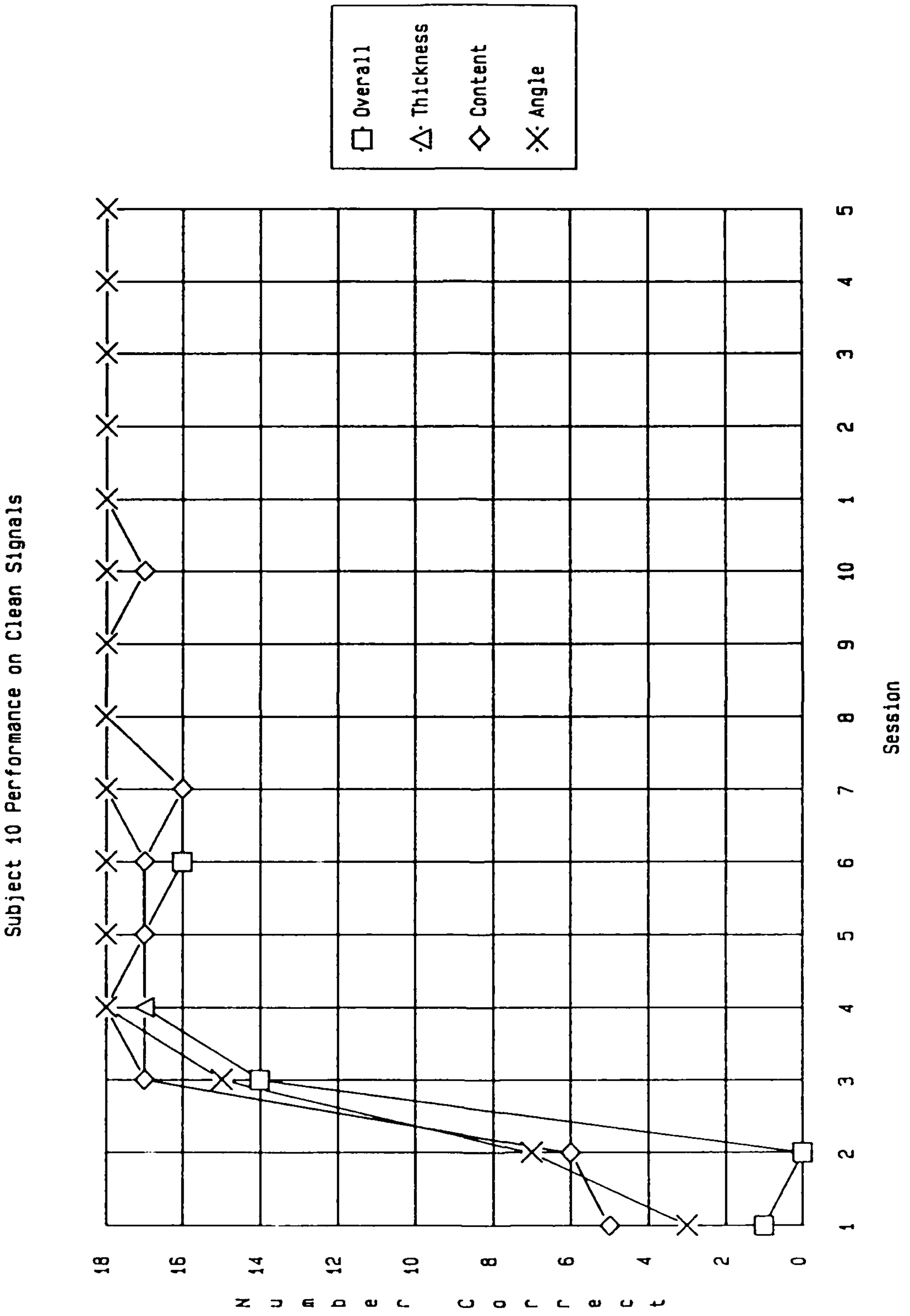


Subject 9 Performance on Clean Signals

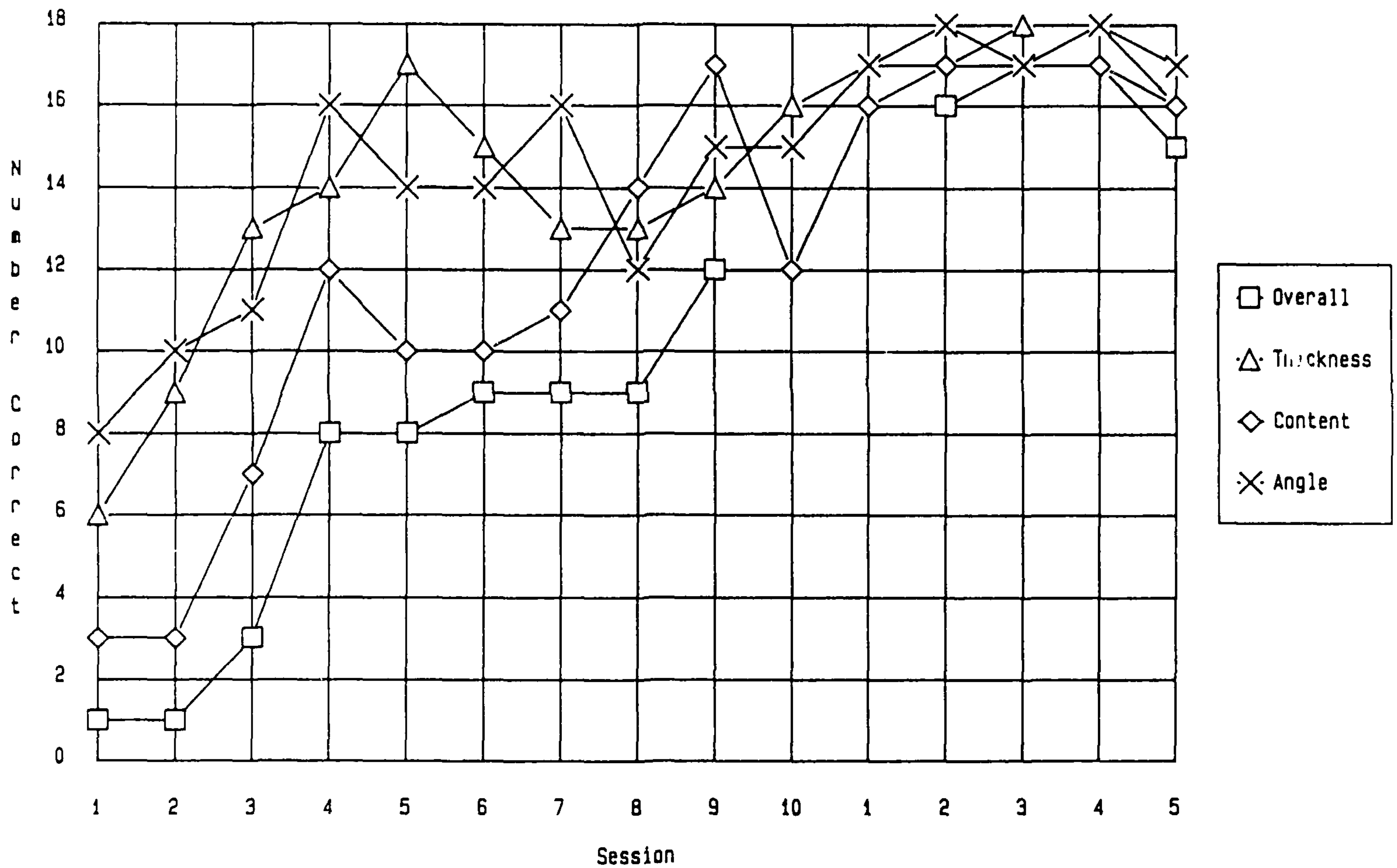


Subject 9 Performance on Noisy Signals




Subject 10 Performance on Clean Signals
Overall
Thickness
Content
Angle
Number Correct
Session
0
2
4
6
8
10
12
14
16
18
1
2
3
4
5
6
7
8
9
10
1
2
3
4
5

Subject 10 Performance on Noisy Signals



ARD

Office of Naval Research Distribution List
Perceptual Science Program

Administrative Contracting Officer
Attn: SCD-C Code S2101A
DCASMA Baltimore
200 Towsontown Blvd., West
Towson, MD 21204-5299

Defense Technical Information Center — 12 copies
Bldg. 5, Cameron Station
Alexandria, VA 22314

Director, Naval Research Laboratory
Attn: Code 2627
Washington, DC 20375-5000

Dr. Dan Greenwood
Netrologic, Inc.
5080 Shoreham Place
Suite 201
San Diego, CA 92122

Dr. Thomas E. Hanna
Naval Medical Research Laboratory
Submarine Systems Department
Naval Submarine Base- New London
Box 900
Groton, CT 06349-5900

Dr. Mien Mann
Code 7033
Naval Air Development Center
Warminster, PA 18974

Office of Naval Research
Code 1114SE Attn: C. G. Lau
800 N. Quincy Street
Arlington, VA 22217-5000

Office of Naval Research
Code 1142BI Attn: T. McKenna
800 N. Quincy Street
Arlington, VA 22217-5000

Office of Naval Research
Code 1142PS Attn: J. J. O'Hare
800 N. Quincy Street
Arlington, VA 22217-5000 — 2 copies

Office of Naval Research
ONR SBIR Administrating Officer
Code 11SP Room 811
800 N. Quincy Street
Arlington, VA 22217-5000

Mr. Paul S. Rau
Code U-33
Naval Surface Weapons Center
White Oak Laboratory
Silver Spring, MD 20903-5000

Dr. Michael Rousseau
Code 2153
Naval Underwater Systems Center
New London Laboratory
New London, CT 06320

Dr. Steven Spiedel
Code 632
Naval Ocean Systems Center
San Diego, CA 92152

Dr. Charles S. Weaver
MAXIM Technologies, Inc.
3000 Patrick Henry Drive
Santa Clara, CA 95054

Dr. Barbara Yoon
DARPA
Information Science & Technology Office
1400 Wilson Blvd.
Arlington, VA 22209